www.sciencemag.org/content/350/6259/434/suppl/DC1

Supplementary Materials for

Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics

Paul N. Evans,* Donovan H. Parks,* Grayson L. Chadwick, Steven J. Robbins, Victoria J. Orphan, Suzanne D. Golding, Gene W. Tyson†

*These authors contributed equally to this work.
†Corresponding author. E-mail: g.tyson@uq.edu.au

Published 23 October 2015, *Science* **350**, 434 (2015)
DOI: 10.1126/science.aac7745

This PDF file includes:

Materials and Methods

Sample collection

Formation water from 11 coalbed methane wells in the Surat Basin (Queensland, Australia) were sampled in May or November of 2013. Biomass was collected by filtering 10 to 50 L of formation water through a 20 µm polypropylene pre-filter and 0.22 µm nitrocellulose collection filter in series using a 142 mm Millipore stainless steel filter housing (#YY3014236, Millipore, MA, USA). The 20 µm and 0.22 µm filters were aseptically folded, placed into separate sterile falcon tubes, and immediately frozen on dry ice for transport back to the laboratory. Well locations, sampling dates, and the depth of formation waters are provided in **Fig. S1**. Geochemical measurements of formation water were collected at the same time as sampling and have been analyzed by Baublys *et al*. (*30*).

DNA extraction and sequencing

While frozen, 4 cm^2 sections of the 20 µm and 0.22 µm filters were cut and placed into triplicate bead tubes containing 0.5 g of 0.1 mm zirconia beads coated with G2 compound (*31*). DNA was extracted using the MoBio PowerSoil DNA Isolation kit (#12888-100, MoBio, CA, USA) after adding 750 µl of PowerMax Soil Power Bead Solution (#12988-10-BS, MoBio, CA, USA). DNA was eluted in 40 µl of distilled water and stored at -80°C, with quantifiable amounts of DNA only produced from the 0.22 µm filters. Libraries were prepared from the 0.22 µm filter extracts using the Illumina Nextera XT DNA Sample Preparation kit and each library sequenced on one-third of an Illumina HiSeq2000 (Illumina, CA, USA) lane to generate 100 bp paired-end reads with a target fragment size of 300 bp. The raw reads have been deposited in the NCBI Sequence Read Archive under accession number PRJNA291107.

Quality control and assembly of sequence data

Metagenomic sequencing generated an average of 4.1 ± 0.6 Gb of paired-end data for each sample. SeqPrep v1.1 (https://github.com/jstjohn/SeqPrep) was used to remove adapters and merge overlapping pairs. Nesoni v0.128 (https://github.com/Victorian-Bioinformatics-Consortium/nesoni) was used to remove homopolymeric reads, trim bases with a Phred quality score <20, and to discard reads ≤30 bp. This quality control resulted in ~37.3 million (±5.5 million) paired-end read and ~2.9 million (±0.67 million) merged or singleton reads per dataset. These reads were independently assembled using the de novo assembly method in CLC Genomics Workbench v7.5 (CLCBio, Denmark) with the word and bubble size determined by the software for each of the 11 metagenomic dataset. For the CX-10 dataset, the selected word and bubble size was 24 and 50, respectively. Scaffolding of the assembled contigs was also performed with CLC Genomics Workbench as part of the assembly processes. Ambiguous bases resulting from the scaffolding were resolved with GapCloser v1.12, part of the SOAPdenovo software suite (*32*), which reduced the number of ambiguous bases by ~50% (8,338,415 to 4,161,643). The resulting 11 assemblies consisted of 57,458 ± 33,326 scaffolds with a length ≥500 bp. The CX-10 assembly consisted of 118,379 scaffolds (191,763 contigs) ≥500 bp with an N50 of 1998 bp, a median length of 829 bp, and a total of ~180 Mbp of unique sequence data.

Recovery and assessment of population genomes

Population genomes were recovered from the CX-10 sample using DBB v1.0.1 (*33*) and MetaBAT v0.24.2 (*34*). Both methods use coverage and tetranucleotide signatures to group scaffolds into putative population genome bins, but use different models for the expected distribution of these statistics which results in the recovery of similar, though not identical, bins. DBB was run with default parameters. Similar to previous studies (*35*), the MetaBAT parameters were set to be very conservative (p1=97, p2=97, minProb=97, minBinned=30) in order to minimize contamination and help distinguish between similar populations. Reads were mapped to assembled scaffolds using BWA v0.7.4 (*36*) with the BWA-MEM algorithm and default parameters. Coverage information for scaffolds was determined independently by DBB and MetaBAT using default parameters and the mapping file produced by BWA. DBB and MetaBAT both recovered two bathyarchaeotal population genomes and the corresponding genomes from each method were combined in order to obtain more complete representatives from this lineage. The BA1 bins recovered with DBB and MetaBAT had 88% of base pairs (74% of scaffolds) in common, whereas the BA2 bins had 89% of base pairs in common (78% of scaffolds). Scaffolds within these combined genome bins with divergent GC-content or tetranucleotide signatures were removed using the *outliers* method of CheckM v1.0.0 (*28*) with default parameters (i.e., scaffolds with genomic features outside the 95th percentile of a typical genome were removed). For the rest of the CX-10 community, we report genomes recovered with DBB, as the two binning methods did not always recover corresponding population genomes. For these genome bins, complementary marker genes were determined with the 'merge' feature of CheckM and bins combined if they had similar GC-content, tetranucleotide signatures, coverage, and placement in the CheckM reference genome tree. Additional scaffolding between contigs within the BA1and BA2 genomes was performed using the 'roundup' mode of FinishM v0.0.2 (https://github.com/wwood/finishm), with GapCloser used to resolve ambiguous bases introduced during this scaffolding procedure.

Estimates of genome completeness and contamination were assessed using CheckM, which uses lineage-specific, single-copy marker genes organized into collocated marker sets in order to provide robust quality estimates (*28*). The BA1, BA2, and E09 SCG were evaluated with an archaeal-specific set of 149 marker genes organized into 107 collocated marker sets. Population genomes estimated to be <20% completeness or >10% contaminated were discarded. Additional genome statistics (e.g., GC-content, coverage, genome size) were also determined with CheckM (**Table S1**). The percentage of reads mapping to recovered population genomes from CX-10 is given in **Table S11**.

Additional assessment of *Bathyarchaeota* genomes

Expected deviations in GC-content and tetranucleotide signatures to the mean value of these statistics was previously calculated for contigs of different lengths over 5,656 trusted reference genomes (*28*). The deviation of scaffolds within the BA1 and BA2 genomes were plotted against these distributions (**Fig. S3, Fig. S4**) using CheckM and scaffolds containing methanogenesis-associated genes checked for consistency with these expected distributions. An analogous plot showing the percentage deviation of each scaffold to the mean coverage of each genome was also produced. A GC-content versus coverage plot (**Fig. S12**) highlighting the relative position of scaffolds within bathyarchaeotal and methanogenic euryarchaeotal population genomes was generated using DBB.

Functional annotation and genome visualization

Assembled contigs were annotated using Prokka v1.11 (*37*) and the RAST annotation server (*38*). Genes with annotations indicating they comprised part of the methanogenesis pathway, along with other critical genes, were further verified using NCBI's BLAST webserver to identify conserved domain (*39*) and through homology search against the Swiss-Prot database (*40*). The operon structure of identified genes was visualized in the Artemis genome viewer (*41*), and metabolic potential examined using KEGG pathways (*42*) derived from the KAAS annotation software (*43*).

Comparison of *Bathyarchaeota* genomes

The amino acid identity between the BA1 and BA2 genomes, along with the E09 single-cell genome (*13*), was measured using a methodology similar to that proposed by Konstantinidis and Tiedje (*15*). Protein-coding sequences (CDS) were identified within all genomes using Prodigal v2.60 (*29*). A pair of genes was considered orthologous if they were the best match to each other as determined using BLASTP v2.2.30+, and both BLAST matches had an identity ≥30% at the amino acid level and an alignment length spanning ≥70% of the length of the query CDS. The CompareM v0.0.5 (https://github.com/dparks1134/CompareM) software package implements this methodology.

Validation and characterization of 16S rRNA sequences

The distribution of reads mapping to the 16S rRNA genes within the BA1 and BA2 genomes was visualized with Tablet (*44*) to ensure paired reads spanned the entire length of these genes (**Fig. S13**). The closest match to the bathyarchaeotal 16S rRNA genes within the May 2013 Greengenes database (*45*), NCBI's nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore), and NCBI's RefSeq database (*46*) were identified using BLASTN v2.2.30+ (**Table S12**).

Coal Oil Point metagenomes

An initial homology search across sequences in the Sequence Read Achieve (SRA; (*47*)) identified several reads with significant sequence similarity to the bathyarchaeotal Surat Basin *mcrA* sequences within two oil polluted marine metagenomes (SRX560108, SRX559946) collected from Coal Oil Point (Santa Barbara, California). Assembled scaffolds for these metagenomes, along with three other metagenomes from Coal Oil Point, were obtained from the Integrated Microbial Genomes database v4.510 (IMG Id: 3300001750, 3300001749, 3300002225, 3300002180, 3300001685; (*48*)). Sequence comparison with BLASTP v2.2.30+ identified five *mcrA* genes among these scaffolds with significant sequence similarity to the bathyarchaeotal Surat Basin *mcrA* sequences (**Fig. 3**; **Table S5**).

Phylogenetic inference

Archaeal genome tree. A set of 195 archaeal genomes previously identified as being of high quality (*28*) were used to establish a set of 144 archaeal-specific marker genes (**Table S13**) suitable for phylogenetic inference as described previously (*49*). Briefly, an initial set of 203 Pfam (*50*) and TIGRFAM (*51*) genes present exactly once in >90% of the high-quality genome set were identified. From this set, 59 were removed from consideration as they were present in multiple copies with divergent phylogenetic histories in >1% of the trusted genomes. The

genome tree was inferred across 295 archaeal genomes obtained from IMG v4.510 (*52*) along with the E09 SCG (*13*) obtained from NCBI (ALXK00000000) and the BA1 and BA2 genomes. Genes were called using Prodigal v2.60 and marker genes identified using HMMER v3.1b1 (http://hmmer.janelia.org) with model specific cut-off values for the Pfam (-cut_gc) and TIGRFAM (-cut_tc) HMMs. Marker genes were aligned with HMMER, concatenated into a single alignment, and phylogenetic inference performed with FastTree v2.1.7 (*53*) under the WAG (*54*) protein substitution model with the gamma approximation used to model rate heterogeneity (*55*).

Support for clades within the genome tree were evaluated using three approaches: non-parametric bootstrapping, marker gene jackknifing, and taxon jackknifing. Bootstrapping support values were calculated from 100 independent bootstrap replicates (*56*). Marker gene jackknifing was performed by removing 50% of the genes at random, while taxon jackknifing was performed by randomly removing 50% of the taxa. One hundred replicates were performed for both jackknifing methods. Support values for marker gene jackknifing can be calculated in the same manner used for non-parametric bootstrap (i.e., the proportion of replicate trees containing a split). Calculating support values for taxon jackknifing is similar, except the absence of taxa within the replicate trees must be taken into account. Specifically, a split $X \cap Y$ in the original tree is considered present in a replicate tree if it contains the split $S = (X \cap T)|(Y \cap T)$, where T is the set of taxa in the replicate tree. Support is calculated over the set of replicates where S is a valid, non-trivial split (*57*).

16S rRNA gene phylogeny. The *Bathyarchaeota* 16S rRNA gene sequences from the BA1 and BA2 genomes along with the E09 SCG (*13*) were uses to identify homologs within GenBank, and the 100 most similar sequences to each of these query sequences downloaded. This dataset was amended to include bathyarchaeotal 16S rRNA genes identified within the Coal Oil Point metagenomes, *Thaumarchaeota* and *Aigarchaeota* reference genomes, and previously described environmental, fosmid, and cosmid sequences (*9, 11*). Sequences were aligned with ssu-align v0.1 (*58*) and the tree inferred with FastTree v2.1.7 using the Generalized Time-Reversible model with the gamma approximation used to model rate heterogeneity. For clarity, clades were organized using the sub-groupings proposed by Meng *et al*. (*9*) and sub-groups collapsed and pruned to focus on the sequences most similar to those obtained from the Surat Basin and Coal Oil Point metagenomes. Support values were determined using 100 non-parametric bootstrap replicates.

McrA sequence phylogeny. The 358 McrA sequences with a length ≥400 aa available at GenBank as of Dec. 18, 2014 were added to the four Surat Basin and five Coal Oil Point McrA proteins ≥400 aa in length (**Table S5**). Identical amino acid sequences were dereplicated to a single sequence which resulted in the removal of 193 of the sequences obtained from GenBank. The set of 174 unique McrA sequences were aligned with MUSCLE v3.8.31 (*59*), and leading and trailing columns represented by <70% of the sequences removed from the alignment which resulted in 710 aligned columns. Genes with amino acids in <70% of the columns were removed, which resulted in the exclusion of 21 genes including one from the Surat Basin and two from Coal Oil Point. ProtTest v3.4 (*60*) identified the WAG and LG (*61*) protein substitution models with observed frequencies (F), invariant sites (I), and the gamma approximation of rate heterogeneity (G) as the two best-fitting models according to the Akaike information criterion. The trimmed MUSCLE alignment was used to infer a McrA amino acid sequence tree with

FastTree v2.1.7 under the WAG+G model (**Fig. 3**). A second tree over the same set of genes was inferred from a multiple sequence alignment obtained with FSA v1.15.9 (*62*) and using RAxML v8.1.11 (*63*) with the LG+F+I+G model as the inference method (**Fig. S14**). Support values for both trees were determined using 100 non-parametric bootstrap replicates.

A McrA tree was also inferred across all McrA sequences identified within the Surat Basin and Coal Oil Point assemblies regardless of their length. Sequences were aligned with MUSCLE, the tree inferred with FastTree under the WAG+G model, and support values determined using 100 non-parametric bootstrap replicates (**Fig. 3**).

Additional Mcr gene trees. Trees for McrB (**Fig. S7**) and McrG (**Fig. S8**) were inferred as follows. For each gene tree, amino acid sequences were obtained from the BA1 and BA2 genomes, and homologs identified across all assembled Surat Basin and Coal Oil Point contigs using BLASTP v2.2.30+. CDS were identified within Surat Basin contigs using Prodigal v2.60 in 'metagenomic' mode. The CDS available at IMG were used from the Coal Oil Point contigs. A gene was considered to be a homolog if it had an amino acid identity ≥30%, an alignment length spanning ≥50% of the length of the query CDS, and an e-value ≤$1e^{-5}$. This expanded set of genes was then used to identify homologs within all reference genomes in IMG v4.510. Homologs were identified across the CDS provided by IMG using the homology criteria specified above. MAFFT v7.221 (*64*) in 'auto' mode was used to align sequences and columns represented by <50% of proteins filtered from the alignment. FastTree under the WAG+G model was used to infer trees and support values determined using 100 non-parametric bootstrap replicates. To aid in comparing these gene trees, a McrA tree was also inferred under these criteria (**Fig. S6**).

Tree editing and visualisation. Trees were visualized and edited using ARB v5.5 (*65*) and further refined using Inkscape v0.48 (inkscape.org).

Comparison of McrA sequence active site conservation and conformation

Coenzyme M, coenzyme B, and cofactor F_{430} binding sites of McrA sequences for representative euryarchaeotal methanogens and methanotrophs were compared to the non-euryarchaeotal sequences from the Surat Basin and Coal Oil Point (**Table S5**). Sequences were aligned with MAFFT using default parameters (**Fig. S9**). The predicted tertiary structure of McrA, McrB, and McrG from the BA1 genomes was predicted using I-TASSER v4.3 with default parameters (*66*). For McrA, McrB, and McrG, the C-score of the top models were 0.68, 2, and 0.45, respectively. Protein-ligand binding sites were verified with the COACH package available as part of the I-TASSER web service (*67*). The presented structural models (**Figs**. **S10**, **S11**) for McrA, McrB and McrG were superimposed on the *Methanopyrus kandleri* crystal structure (PDB ID: 1E6V) as this structure resulted in the best fit to the BA1 McrA model. For McrA, McrB, and McrG, the TM-score for the alignment to the 1E6V structure were 0.897, 0.938, and 0.919, respectively. The RMSD between the TM-aligned residues were 0.54, 0.79, and 0.64 angstroms, respectively.

Environmental distribution of *Bathyarchaeota* McrA sequences

DIAMOND v0.7.3 was used to identify reads within publicly available metagenomes with higher sequence similar to a novel Surat Basin or Coal Oil Point McrA amino acid sequence (**Table S5**) than to any of the 85 euryarchaeotal McrA sequences available in IMG v4.510.

Reads from 2,705 metagenomes of non-human origin within the SRA were compared using the BLASTX method of DIAMOND. The 57 genomic fragments with a best match to either a Surat Basin or Coal Oil Point McrA protein, and significant sequence similarity (e-value > $1e^{-5}$, bit score > 50) to this best matching protein were reported (**Table S6**). Homology search against the GenBank protein database using BLASTX resulted in 30 of the 57 reads being identified as highly divergent from an annotated euryarchaeotal McrA protein, while the other 27 fragments had no significant match within the GenBank database.

16S rRNA reads belonging to the *Bathyarchaeota* were identified in the SRA metagenomes by mapping reads to the 99% dereplicated version of the Greengenes database (May, 2013) using BWA with default parameters. Reads were considered properly mapped if they were a primary alignment, ≥50% of the read aligned to the reference sequence, and the percentage identity of the alignment was ≥75%. A read aligned to any of the 442 16S rRNA gene sequences in Greengenes marked as being from the class MCG were classified as belonging to the *Bathyarchaeota* (**Table S7**).

Supplementary Text

Presence of *mcrCD* genes putatively associated with BA1

All known methanogens, anaerobic methanotrophs, and the BA2 genome have *mcrCD* genes in addition to *mcrABG*. This suggests that the BA1 genome likely contains *mcrCD* genes in addition to the identified *mcrABG* operon. Within the BA2 genome the *mcrCD* genes sit adjacent to each other, but 11 genes downstream from the *mcrA* gene (~10 kb) indicating these genes are not always collocated within the *Bathyarchaeota*. The *mcrABG* genes within BA1 reside on a contig that is only 9.3 kb in length and are located near the end of the contig. As such, it is unsurprising that the *mcrCD* genes do not reside on this contig and suggests they may reside on an unbinned contig.

We have identified an unbinned contig (~12 kb) assembled from the CX-10 metagenome that contains *mcrCD* genes and appears to be bathyarchaeotal in origin based on 5 of its 13 genes, including the *mcrCD* genes, having higher sequence similarity to genes located on the BA2 contig containing *mcrABCDG* than to any gene in a IMG v4.510 reference genome. In addition, these 5 genes fall in the same relative order on the two contigs. Unfortunately, we were unable to confidently link this contig with the BA1 genomes as the contig has a tetranucleotide signature slightly outside the expected distribution of a typical genome. Additional efforts to associate this contig with the BA1 genome have also been unsuccessful (e.g., paired reads linking this contig to a contig in the BA1 genome; manual examination of the de Brujin assembly graph). Nonetheless, given the expectation of *mcrCD* genes in the BA1 genome, the sequence similarity and synteny shared between this contig and BA2 homologs, and the high variance of tetranucleotide signatures when calculated over relatively short contigs, we believe it is likely, though not rigorously established, that this contig comprises part of the BA1 genome.

Proposed mechanism for methanogenesis in BA1 and BA2

In the BA1 genome (**Fig. 2**), CO_2 fixation could occur via the Wood-Ljundahl pathway, where carbon is assimilated through C1 carriers, methanofuran and tetrahydromethanopterin (H_4MPT), CO dehydrogenase/acetyl-CoA synthase (*cdhABCDE*) and pyruvate oxidoreductase (*porABDG*) leading to the production of pyruvate in a manner typical of archaea (*68*). The formylmethanofuran dehydrogenase (*fwdABCD*) would reduce the CO_2 forming formylmethanofuran using reduced ferredoxin. After being transferred from methanofuran to H_4MPT by formylmethanofuran transferase (*ftr*), stepwise reduction to methyl-H_4MPT would occur via methenyl-H_4MPT cyclohydrolase (*mch*), methylene-H_4MPT dehydrogenase (*mtd*) and methylene-H_4MPT reductase (*mer*). Electrons for the reduction of the H_4MPT bound C1 compound would be gained by H_2 oxidation by the F_{420}-reducing hydrogenase (*frhABG*) to reduce the C1 compounds via coenzyme F_{420}. H_2 would likely be generated by reoxidation of reduced ferredoxin by a putative membrane bound energy converting hydrogenase (*echABCDEF*).

Alternatively, acetyl-CoA formed during CO_2 assimilation could be diverted by an ADP-forming acetyl-CoA synthase (*acd*) to produce acetate and ATP. The presence of a CO dehydrogenase (*cooS*) in the BA1 genome also suggests that CO may be used as a carbon source, using a mechanism similar to *Methanosarcina* and *Methanobacterium* species (*69–71*). Methanol or other methylated compounds may be assimilated via methyl-H_4MPT by methyl-H_4MPT methyltransferase (*mtrH*) and an associated corrinoid protein, similar to the mechanism hypothesized by Harms and Thauer (*18*). It is also plausible that *mtrH* and the associated

corrinoid protein may donate unknown methyl-groups to coenzyme M as a substrate for methanogenesis.

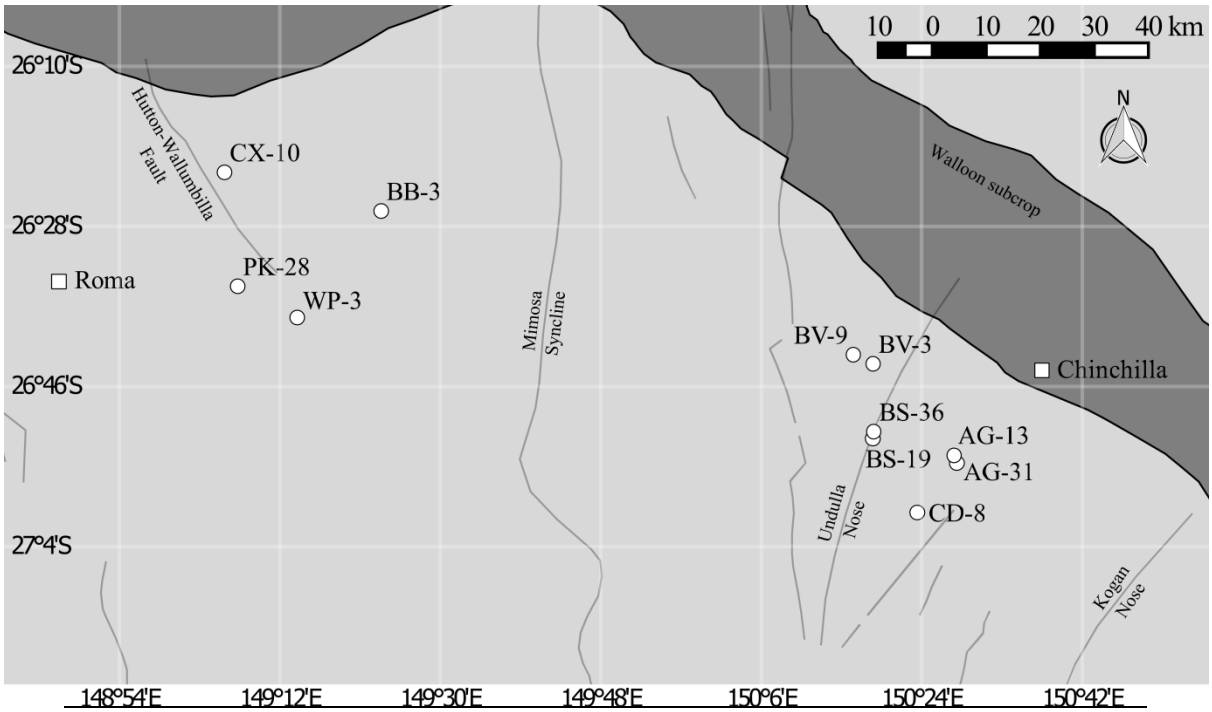
Methanogenesis catalyzed by the methyl-coenzyme M reductase complex (McrABG) results in methane and a coenzyme M (CoM) and coenzyme B (CoB) disulfide (CoM-S-S-CoB). This CoM-S-S-CoB disulfide is reduced by either heterodisulfide reductases *hdrABC* or *hdrD* to regenerate these coenzymes in an oxidation/reduction cycle. In BA1, *hdrABC* is collocated with a F_{420}-non-reducing hydrogenase (*mvhADG*) genes typically seen in electron bifurcating complex of euryarchaeotal methanogens (*2*). If reduced ferredoxin is made through electron bifurcation it could be reoxidized at the Ech complex, providing BA1 with a transmembrane potential and an additional H_2 which could be used in the next heterodisulfide reduction reaction, as has been proposed for members of the *Methanomassiliicoccales* (*72*). Alternatively, CoM and CoB could be regenerated by *hdrD*. The three *hdrD* copies are collocated with *glcD*-like genes in a configuration similar to the lactate dehydrogenase operon (*glpC* and *glcD*) in *Archaeoglobus fulgidus* (*19*), suggesting that lactate may be a source of electrons to reduce the CoM-S-S-CoB disulfide. This mechanism is similar to that of thiol:fumarate reductase (*tfr*) (*73*), which is not linked to energy generation.

Like BA1, BA2 (**Fig. 2**) has the machinery for the CoM-S-S-CoB cycling with *mcrABG*, *hdrABC/mvhADG,* and *hdrD*/*glcD* homologs. However, the absence of recognized methyltransferases suggests other mechanisms for the transfer of methyl groups, potentially by homologs similar to the 12 divergent methyltransferases that were observed in the BA1 genome. Also, the absence of many genes comprising the Wood-Ljundahl pathway in the BA2 genome suggest that acetate is used as a carbon source rather than CO or CO_2 based on the presence of an acetyl-CoA forming synthase (*acs*), which is absent in the BA1 genome. Neither genome has acetate kinase (*ack*) or phosphotransferase (*pta*).

Proposed mechanism for methane oxidation in BA1 and BA2

It has been proposed that members of the *Bathyarchaeota* could be responsible for methane oxidation in sulfate methane transition zones of marine sediments (*14*). Methane oxidation coupled to sulfate reduction is thought to occur in some marine sediments through a syntrophic interaction between euryarchaeotal methanotrophs and bacterial sulfate reducers. In this model, the traditional methanogenesis pathway is reversed to oxidize methane, and electrons produced through this pathway are shared with the sulfate reducing partner. Very little free energy is available to this reaction, calculated to be between -18 and -35 kJ/mol methane in various marine habitats (*3*). Euryarchaeotal methanotrophs appear to have sophisticated membrane bound electron transport chains with energy conservation strategies that rely on proton or sodium membrane potentials and ATPases (*25*, *74*). BA1 and BA2 on the other hand do not appear to have ATPases or membrane-bound electron transport, so any methane oxidation strategy would have to rely on soluble cytoplasmic strategies for energy conservation. With these constraints in mind the only possible methane oxidation strategy would seem to be methane oxidation with CO_2 as the electron acceptor, i.e. reverse acetoclastic methanogenesis. Acetoclastic methanogenesis has been calculated to proceed with extremely small free energies ~-10kJ/mol (*75*). Reversing this reaction would result in small positive free energies, which could potentially become negative and spontaneous under conditions with sufficiently low concentrations of the end product acetate. These low acetate concentrations could be achieved through a syntrophic interaction between the *Bathyarchaeota* and an acetate-utilizing sulfate reducing bacteria. Methane would be oxidized to the level of methyl-tetrahydromethanopterin (methyl-H4MPT),

with the electrons being used to reduce one CO_2 to [CO] in the CODH/ACS complex. Acetyl-CoA made from methyl and [CO] could then be used for ATP generation by Acd, producing acetate that could be transported to a sulfate reducing partner. In this scheme, the bathyarchaeotal reaction would be: $CH_4 + HCO_3^- \rightarrow H_3CCO_2^- + H_2O$, and the reaction performed by the sulfate reducer would be: $H_3CCO_2^- + SO_4^{2-} \rightarrow 2HCO_3^- + HS^-$, with the combined reaction being the same as that proposed for euryarchaeotal methanotrophs and sulfate reducing bacteria: $CH_4 + SO_4^{2-} \rightarrow HCO_3^- + HS^- + H_2O$. Since acetate would only be produced as a transient intermediate its concentration could be optimized such that the small free energy associated with the overall reaction would be split evenly between the two partners.



Site id	Site name	Sampling date	Latitude	Longitude	Formation water interval (m)
AG-13	Argyle 13	11/27/2013	-26.9134	150.4656	261 to 615
AG-31	Argyle 31	11/27/2013	-26.8991	150.4606	265 to 604
BB-3	Ben Bow 3	11/25/2013	-26.4411	149.3893	515 to 835
BV-3	Bellevue 3	11/28/2013	-26.7275	150.309	172 to 480
BV-9	Bellevue 9	11/28/2013	-26.7104	150.272	189 to 527
BS-19	Berwyndale South 19	11/27/2013	-26.8675	150.3086	362 to 566
BS-36	Berwyndale South 36	11/27/2013	-26.8546	150.3103	309 to 617
CD-8	Codie 8	5/8/2013	-27.0061	150.3916	460 to 804
CX-10	Coxon Creek 10	11/25/2013	-26.3686	149.0963	90 to 442
PK-28	Pickanjinnie 28	11/25/2013	-26.5821	149.1209	610 to 821
WP-3	Washpool Creek 3	11/25/2013	-26.6407	149.2323	569 to 978

Fig. S1.

Formation waters sampled within the Surat Basin. Sampling sites are shown as circles and nearby towns as squares. The full name of each sampling well along with the sampling date, geographic location, and depth interval of formation waters are provided in the table.

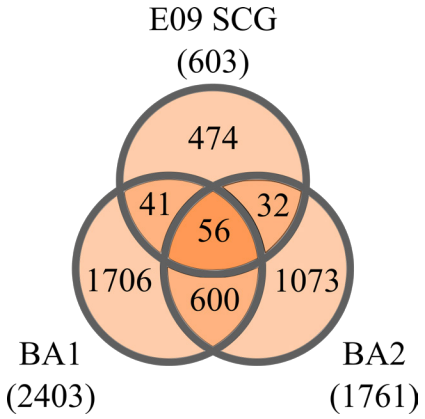


Fig. S2.

Shared gene content between representatives from the archaeal phylum *Bathyarchaeota*. The total number of protein coding genes in each genome is shown in parentheses.

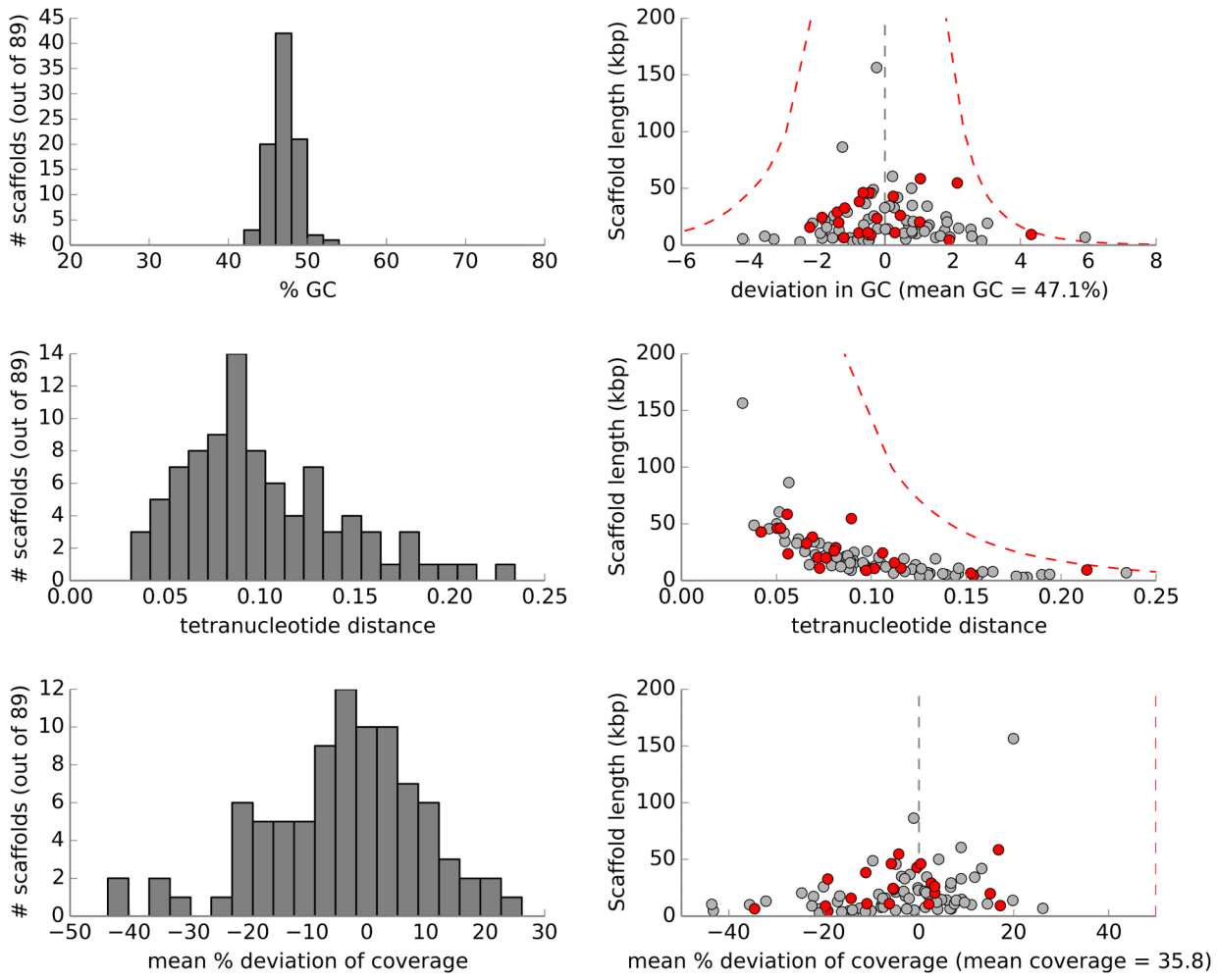


Fig. S3.

Evaluation of the statistical properties of scaffolds comprising the BA1 genome. The histograms show the GC-content, tetranucleotide distances (TD) from the mean tetranucleotide signature of the genome, and coverage of all scaffolds within the genome. Each point in the scatterplots represents a scaffold plotted as a function of its length (y-axis) and deviation from the mean GC-content, tetranucleotide signature, or coverage of the genome (x-axis). The dashed red lines represent the 95th percentile for each of these statistics as determined by fragmenting finished IMG genomes into non-overlapping windows of a fixed size. Scaffolds in red contain one or more of the methane metabolism-associated gene specified in **Table S4**.

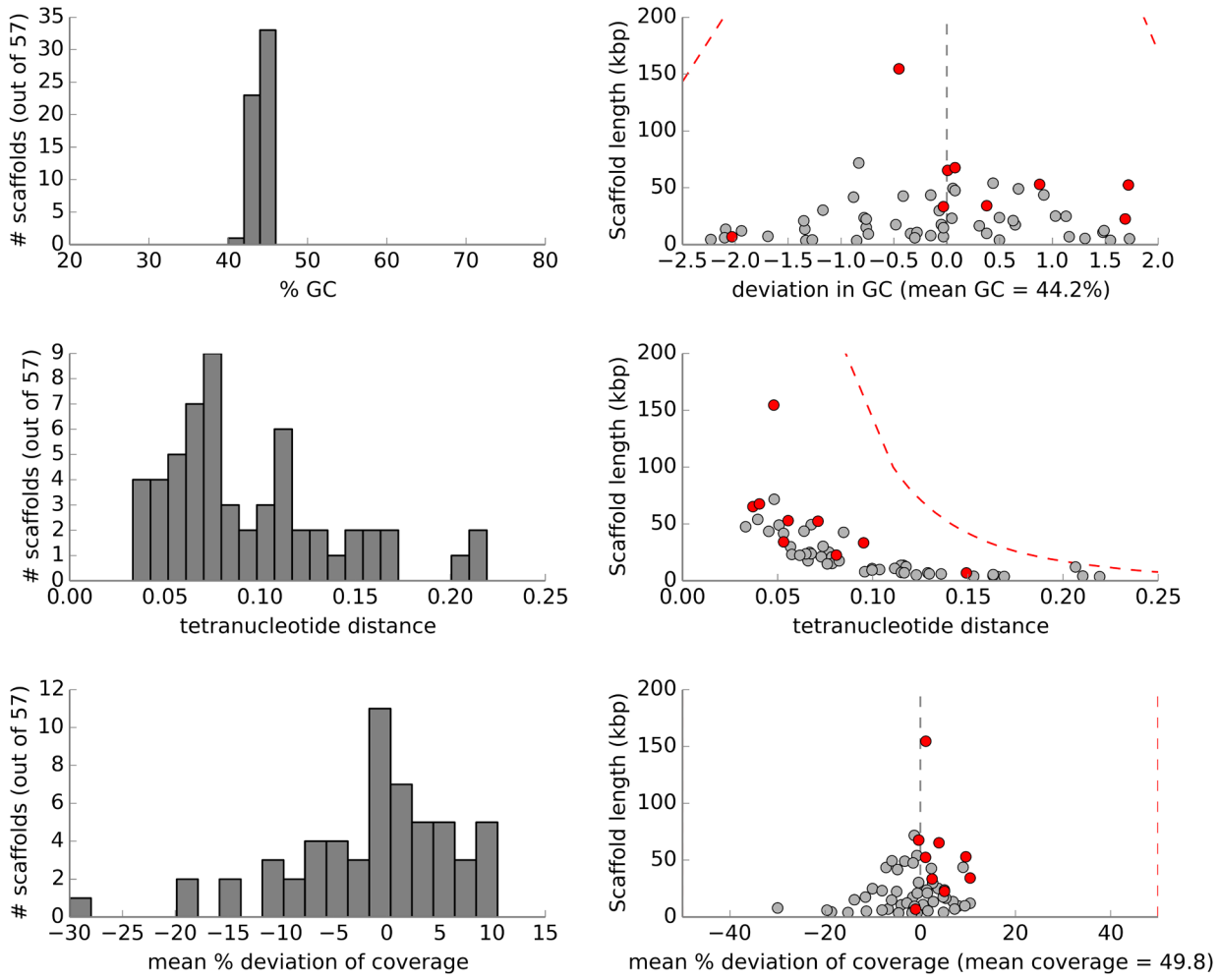


Fig. S4.

Evaluation of the statistical properties of scaffolds comprising the BA2 genome. The histograms show the GC-content, tetranucleotide distances (TD) from the mean tetranucleotide signature of the genome, and coverage of all scaffolds within the genome. Each point in the scatterplots represents a scaffold plotted as a function of its length (y-axis) and deviation from the mean GC-content, tetranucleotide signature, or coverage of the genome (x-axis). The dashed red lines represent the 95th percentile for each of these statistics as determined by fragmenting finished IMG genomes into non-overlapping windows of a fixed size. Scaffolds in red contain one or more of the methane metabolism-associated gene specified in **Table S4**.

(A) **ME1/ME2 primer set of Hales *et al.* (*76*)**

	ME1 GCMATGCARATHGGWATGTC	Mismatches
M_smithii	GTAAGGGCATGTGACGGAGGTACTGTATCCAGATGGTCTGCTATGCAGATTGGTATGAGT	3
M_paldcla	GTCAGGACCTGCGACGGCGGCACGACCTCCAGATGGTCGGCCATGCAGATTGGCATGTCG	1
M_cncili	GTCAGGACCTGCGACGGCGGCACCACATCCAGATGGAGCGCCATGCAGATGTGTATGACC	3
M_hngti	TCCCGTACATGTGATGGTGGAACTACCTCCCGGTGGTCAGCCATGCAGATTGGTATGTCC	0
M_holnda	AGCAGGGTAACAGACGGTGCCCAGACTACCAGATGGATGGCCATGCAGGTTGGTATGTCC	1
M_jannshi	GGTAGAGTTTGTGATGGGGGAACAATAGCAAGATGGAGTGCTATGCAGATTGGAATGTCA	1
M_inststns	GGACGTGTGTGCGATGGTGGTACAATGTCCAGATGGAGCGCCATGCAAATTTCAATGTCT	2
M_thrmoauto	TCCCGAACTTGTGACGGTGGTACCACCTCACGATGGTCAGCTATGCAAGTGGGTATGTCA	3
M_stdtm	GTACGTGCTTGTGATGGAGGTACAACCTCCAGATGGTCTGCTATGCAAATTGGTATGAGT	3
M_barkeri	TCCAGAACAACAGACGGTGCTCAGACCTCCAGGTGGGCAGCCATGCAGATCGGTATGTCT	0
M_kandleri	GGCCACGTCTGTGACGGAGCGACCATGTACCGATGGGCCGCGATGCAGATCGCGATGTCG	3
BA1 24_9	ATAAGGATGGCGGATGGGGAAGAGGCCGGGAGGTGGAGCGCTATGCAGAACACGATGGCC	6
BA2 13_151	ATAAGGATGGCTGATGGTGACGAGGCTCATAGGTGGGCTGCTATGCAGAACACGATGGCT	6
BV3 3440_1	----------	
CD8 186481_1	----------	
BS36 5099_1	ATAAGAACGGCGGATGGGGAAGAGGCCCACCGATGGGCGGCTATGCAGAATTCGATGGCT	6
BS36 20936_1	----------	
AG31 16992_1	----------GCT	NA
BS19 9145_2	ATCCGAATGGCGGATGGAGACGTAGCGGCGAGATGGGCCGCAATGCAGAATACGATGGCT	5
BB3 1324_1	ATAAGGACGGCAGATGGAGATGAAGCCCACAGGTGGGCAGCCATGCAGAATACGATGGCT	5
BB3 5336_3	----------	
BV9 1718_1	----------	
BV3 6795_1	ATCCGAATGGCGGATGGAGACGTAGCGGCGAGATGGGCCGCAATGCAGAATACGATGGCT	5
CX10 14344_1	ATTAGAATGGCGGATGGAGACATAGCAACTAGATGGGCTGCTATGCAGAACACGATGGCA	6
CX10 1922_1	ATAAGGACGGCAGACGGAGATGAAGCGCACAGGTGGGCAGCTATGCAGAATACAATGGCT	5
COP3-100133354	GTGCGGATGGCGGATGGAGAAGCTGCTGGAAGATGGGCCGCGATGCAAAATACGATGGCG	6
COP3-100174111	GTCAGGATGATGGATGGTGGTGCAATCGTTAGAAGTGCAGCACAGGGGGCAACGATGGCA	10
COP3-100202833	GTCAGGATGATGGACGGTGGAGCAATAGTCAGAAGTGCAGCACAGTCTGCAACGATGGCC	11
COP3-100352853	GTCCGTAACATGGACGGTGCCGCAGTTGTAAGAAGCGCTGCACAAGCAACGACGATGGCT	11
COP3-100679051	GTTAGGAATATGGATGGCGCTGCAATCGTGAGAAGTGCCGCTCAATCTACAACGATGGCT	12

	ME2 ACTAYCCHAACTAYGCMATGA (reverse complemented)	Mismatches
M_smithii	AGAGGACCAAACTACCCGAACTATGCTATGAATGTAGGTCACCAGCCAGAATATGCAGGT	2
M_paldcla	AGAGGCCCGAACTACCCGAACTACGCGATGAACGTCGGCCACCAGGGTGGCTACGCCGGT	2
M_cncili	AGAGGCGCCAACTACCCGAACTACGCCATGAACGTGGGCCACCAGGGCGAGTACGCAGGC	1
M_hngti	CGTGGTCCAAACTACCCGAACTACGCAATGAACGTGGGTCACCAGGGAGAATACGCAGCA	1
M_holnda	CGTGGACCAAACTACCCCAACTACGCCATGAACGTAGGTCACCAGGGTGGATACGCTGCA	0
M_jannshi	AGAGGGCCTAACTATCCAAACTACGCAATGAACGTTGGTCACCAAGGAGAATATGCTGGA	0
M_inststns	AGAGGACCCAACTATCCAAACTACGCAATGAACGTTGGTCACCTTTGCGGTTACGCTGGT	0
M_thrmoauto	AGAGGAGCTAACTACCCCAACTACGCCATGAATGTGGGTCACCAGGGTGAATATGCAGGT	0
M_stdtm	AGAGGTCCTAACTACCCTAACTACGCAATGAACGTAGGTCACCAACCTGAATATGCAGGT	0
M_barkeri	CGTGGTCCAAACTACCCCAACTACGCAATGAACGTTGGTCACCAGGGTGGATACGCAGGT	0
M_kandleri	CGTGGTCCGAACTACCCGAACTACGCGATGAACGTGGGTCACCTGGGCGAGTACGCGGGT	2
BA1 24_9	AGGGGCTACAACACCCCATACCAATCTTTCACGGCTGGGCACGGCGTCATCCGGGAGGTG	10
BA2 13_151	AGAGGTCTTAACTACCCGATGCAATCTTTCAGCGCAGCCCATGGTTCTCTAAGGGATGCA	10
BV3 3440_1	AGAGGCCTTAACTACCCGATGCAAACTTTCACAGCAGCCCACGGTGTCCTC----------	10
CD8 186481_1	AGAGGTCTTAACTATCCGATGCAATCTTTCAGTGCAGCGCACGGTGCTATAAGGGATACG	10
BS36 5099_1	----------	
BS36 20936_1	AGAGGTCTTAACTATCCGATGCAATCTTTCAGTGCAGCGCACGGTGCTATAAGGGATACG	10
AG31 16992_1	CGTGGGATGAACTATCCAATGCACGCCTTCACGCCAGCGCACGCAGCTATGAGGCAAGCT	6
BS19 9145_2	AGAGGTCTTAACTATCCGATGCAATCTTTCAGTGCAGCGCACGGTGCTATAAGGGATACG	10
BB3 1324_1	----------	
BB3 5336_3	AGAGGCCTTAACTACCCGATGCAAACTTTCACGGCAGCCCACGGTGTCCTCCGAGATACG	10
BV9 1718_1	AGAGGTCTTAACTATCCGATGCAATCTTTCAGTGCAGCGCACGGTGCTATAAGGGATACG	10
BV3 6795_1	----------	
CX10 14344_1	AGGGGTCTAAATTACCCTATGCAATCGTTCAGCGCGGCTCATGGTGCTCTTAGGGCTGCG	10
CX10 1922_1	----------	
COP3-100133354	----------	
COP3-100174111	AAAGGTCCGAATGTTCCATACATGTCATATACAGTCGGTGCACTTGGGTTCCCGATGTTC	11
COP3-100202833	AAAGGTCCGAATGTACCATACATGTCGTACACTGTTGGCGCTCTTGGGATGCCGATGTTC	13
COP3-100352853	AAAGGAGAGAATGTGCCATACATGTCCTATACCGCAGTAGGTGGTAGCCTTGGTGTAGGA	12
COP3-100679051	----------	

(B) **MCRf/MCRr primer set of Springer *et al.* (*77*)**

```
                                                  MCRf      TAYGAYCARATHTGGYT   Mismatches
M_smithii          ------------------GTTATTGCAGGTGCTGCTGCGCTATATGATCAGGTTTGGTTA      1
M_paldcla          ------------------GTAGTCGCGGCCGGCTGCATGCTGTATGACCAGATCTGGCTC      0
M_cncili           ------------------GTGGTCGCCGCAGGCTCCATGTTCTATGATCAGATCTGGCTG      0
M_hngti            ------------------GTTGTCGGTGCAGGTTGTATGCTCTTCGACCAGATCTGGCTC      1
M_holnda           ------------------GTAGTCGGTGCAGGCTGTATGCTCTACGACCAGATCTGGCTT      0
M_jannshi          ------------------GTTGTTGCTGCTGGGGCTATGTTGTATGACCAAATCTGGTTA      0
M_instststns       ------------------ACTGTAGCTCTCGGTGCAATCATCTTCGATCAGATTTACCTT      3
M_thrmoauto        ------------------GTCATCGCAGGTGCAGCAGCACTCTACGACCAGGTCTGGCTC      1
M_stdtm            ------------------GTTATTGGTGGAGCAGCACCATTATACGACCAAGTATGGTTA      1
M_barkeri          ------------------GTCGTCGGTGCAGGCTGTATGCTCTACGACCAGATCTGGCTC      0
M_kandleri         ------------------GTGATCGCCGCCGGAGCGATGCTCTACGACCAGATCTGGCTA      0
BA1 24_9           CACGCCATGTCTGAA---GGCATCTGCATCCCAGCGATCATAGACAGCGCCTATTGGTTC      8
BA2 13_151         CGTGCTATGTACGAAGGTATCGGCGCAATCGCAGCAGTTCTTGACAATATGATATGGCTG      4
BV3 3440_1         ---------GGCGAGGGTTCGGGCGGGATAGCAGCGATTCTTGATAATTCATTATGGCTT      5
CD8 186481_1       ----------------------GCATCGATTGGAGCAATTCTTGACAACACGATAGGCTTA      6
BS36 5099_1        CTTGCGGCAAGCGAGGGTTCGGGTGGGATAGCAGCTATTCTTGATAATGCACTATGGCTT      5
BS36 20936_1       CACGCCTTGAT---TGACTCTGCATCGATTGGAGCAATTCTTGACAACACGATATGGCTA      4
AG31 16992_1       CGCGTTATCATTGAAGGAATAGGTGGGATATCCTCTATGATTGAAAATACATTGTGGTTT      7
BS19 9145_2        CGCGCCTTAAT---CGACTCCGCATCGATTGGAGCAATCCTTGACAATACAATATGGCTA      4
BB3 1324_1         CTTGCGCTGGGCGAGGGTTCCGGTGGGATAGCAGCAATTCTTGATAATTCATTATGGCTG      5
BB3 5336_3         CTTGCAATAGGCGAGGGTTCAGGCGGGATAGCAGCAATTCTTGACAATTCATTATGGCTT      5
BV9 1718_1         CACGCCTTGAT---TGACTCTGCATCGATTGGAGCAATTCTTGACAACACGATATGGCTA      4
BV3 6795_1         ------------------------------------------------------------
CX10 14344_1       AAAGCTATAGTCGAAGGTGCTGGTGGGATAGGCGCGGTAATTGATAACTCGGTATGGCTT      5
CX10 1922_1        CTTGCATTAGGCGAGGGTTCAGGTGGGATAGCAGCAATTCTTGACAATTCATTATGGCTT      5
COP3-100133354     AGTAGGCTGAACGAGGGAGTTGGTGGGATATCCGCCATGCTCGATAATACGTTGTGGTTG      6
COP3-100174111     ATGGGCGCAGTTGGTAGTGGCGGTGGAATGTGGGTAGAATTACTCACAAATCTGTGGTAT     10
COP3-100202833     TTCGGAGGTATTGCCGCGGGTGGTCCGATGTGGACAGAACTTCTGACGAACTGCTGGTAC     11
COP3-100352853     CTCAACGCCCTGATGGGTTGTGGGCAGATGTGGGCGGAAATGGTAGAGAACATATGGTTA      7
COP3-100679051     ATGCAGGCAATGATGCAGTGTGCAGCCATGAGTTCGGAACTTATAGAAATATATGGTAT      8

                                MCRr   AAYTAYGCNATGAAYGT  (reverse complemented)   Mismatches
M_smithii          AGAGGACCAAACTATCCGAACTATGCTATGAATGTAGGTCACCAGCCAGAATATGCAGGT      0
M_paldcla          AGAGGCCCGAACTACCCGAACTACGCGATGAACGTCGGCCACCAGGGTGGCTACGCCGGT      0
M_cncili           AGAGGCGCCAACTACCCGAACTACGCCATGAACGTGGGCCACCAGGGCGAGTACGCAGGC      0
M_hngti            CGTGGTCCAAACTACCCGAACTACGCAATGAACGTGGGTCACCAGGGAGAATACGCAGCA      0
M_holnda           CGTGGACCAAACTACCCCAACTACGCCATGAACGTAGGTCACCAGGGTGGATACGCTGCA      0
M_jannshi          AGAGGGCCTAACTATCCAAACTACGCAATGAACGTTGGTCACCAAGGAGAATATGCTGGA      0
M_instststns       AGAGGACCCAACTATCCAAACTATGCAATGAACGTTGGTCACCTTTGCGGTTACGCTGGT      0
M_thrmoauto        AGAGGAGCTAACTACCCCAACTACGCCATGAATGTGGGTCACCAGGGTGAATATGCAGGT      0
M_stdtm            AGAGGTCCTAACTACCCTAACTACGCAATGAACGTAGGTCACCAACCTGAATATGCAGGT      0
M_barkeri          CGTGGTCCAAACTACCCCAACTACGCAATGAACGTTGGTCACCAGGGTGGATACGCAGGT      0
M_kandleri         CGTGGTCCGAACTACCCGAACTACGCGATGAACGTGGGTCACCTGGGCGAGTACGCGGGT      0
BA1 24_9           AGGGGCTACAACACCCCATACGCATCTTTCACGGCTGGGCACGGCGTCATCCGGGAGGTG      9
BA2 13_151         AGAGGTCTTAACTACCCGATGCAATCATTCAGCGCAGCCCATGGTTCTCTAAGGGATGCA      9
BV3 3440_1         AGAGGCCTTAACTACCCGATGCAAACTTTCACAGCAGCCCACGGTGTCCTC---------      10
CD8 186481_1       AGAGGTCTTAACTATCCGATGCAATCTTTCAGTGCAGCGCACGGTGCTATAAGGGATACG      9
BS36 5099_1        ------------------------------------------------------------
BS36 20936_1       AGAGGTCTTAACTATCCGATGCAATCTTTTAGTGCAGCGCACGGTGCTATAAGGGATACG      9
AG31 16992_1       CGTGGGATGAACTATCCAATGCACGCCTTCACGCCAGCGCACGCAGCTATGAGGCAAGCT      9
BS19 9145_2        AGAGGTCTTAACTATCCGATGCAATCTTTTAGTGCAGCGCACGGTGCTATAAGGGATACG      9
BB3 1324_1         ------------------------------------------------------------
BB3 5336_3         AGAGGCCTTAACTACCCGATGCAAACTTTCACGGCAGCCCACGGTGTCCTCCGAGATACG     10
BV9 1718_1         AGAGGTCTTAACTATCCGATGCAATCTTTTAGTGCAGCGCACGGTGCTATAAGGGATACG      9
BV3 6795_1         ------------------------------------------------------------
CX10 14344_1       AGGGGTCTAAATTACCCTATGCAATCGTTCAGCGCGGCTCATGGTGCTCTTAGGGCTGCG      9
CX10 1922_1        ------------------------------------------------------------
COP3-100133354     ------------------------------------------------------------
COP3-100174111     AAAGGTCCGAATGTTCCATACATGTCATATACAGTCGGTGCACTTGGGTTCCCGATGTTC      9
COP3-100202833     AAAGGTCCGAATGTACCATACATGTCGTACACTGTTGGCGCTCTTGGGATGCCGATGTTC      8
COP3-100352853     AAAGGAGAGAATGTGCCATACATGTCCTATACCGCAGTAGGTGGTAGCCTTGGTGTAGGA      9
COP3-100679051     ------------------------------------------------------------
```

(C) **MLf/MLr primer set of Luton *et al.*** (*78*)

	MLf GGTGGTGTMGGATTCACACARTAYGCWACAGC	Mismatches
M_smithii	TGGTTAGGTTCTTACATGTCAGGAGGTGTAGGTTTCACTCAGTATGCAAGTGCAACTTAT	5
M_paldcla	TGGCTCGGCTCCTACATGTCCGGTGGTGTCGGCTTCACGCAGTATGCGACCGCCGCCTAC	4
M_cncili	TGGCTGGGAAGCTACATGTCCGGCGGTGTCGGTTTCACCCAGTATGCTACTGCTGCCTAC	4
M_hngti	TGGCTCGGTTCCTACATGTCTGGTGGTGTCGGATTTACCCAGTATGCAACCGCAGCATAC	3
M_holnda	TGGCTTGGATCATACATGTCCGGTGGTGTGGGTTTCACACAGTACGCAACAGCTGCATAC	2
M_jannshi	TGGTTAGGAGGATACATGTCTGGTGGTGTCGGATTTACACAGTATGCTACAGCAACCTAT	1
M_instststns	TACCTTGGATCATACATGTCTGGTGGTGTCGGATTCACACAGTACGCTACTGCAGCATAC	1
M_thrmoauto	TGGCTAGGTTCCTACATGTCAGGTGGTGTAGGATTCACCCAGTACGCTACTGCAGCTTAC	2
M_stdtm	TGGTTAGGTTCATACATGTCTGGTGGTGTCGGATTCACACAATATGCTACAGCAGCATAC	0
M_barkeri	TGGCTCGGATCCTACATGTCCGGTGGTGTCGGATTCACACAGTATGCAACAGCTGCATAC	0
M_kandleri	TGGCTAGGATCCTACATGTCAGGAGGTGTCGGTTTCACGCAGTACGCGACGGCCGTGTAC	5
BA1 24_9	TGGTTCGGGTTTTACATGAGCGGCGGAATAGGGTTCACGAACACCACTGCGGGGGCAGCC	13
BA2 13_151	TGGCTGGGATTCTACATGAGTGGCGGTATAGGTTTCTCAAACACGGTATCCGCTGCAGGT	12
BV3 3440_1	TGGCTTGGATTTTACATGAGCGGTGGTATAGGTTTCTCAAATACCGTAGCCTCAGCAGCT	11
CD8 186481_1	TGGCTAGGGTTCTATATGAGCGGTGGTATAGGGTTCTCAAACACAGTGGGCGGAGCGGCC	14
BS36 5099_1	TGGCTTGGATTCTACATGAGCGGCGGTATAGGTTTCTCAAATACCGTAGCCTCAGCAGCT	12
BS36 20936_1	TGGCTAGGGTTCTATATGAGCGGTGGTATAGGGTTCTCAAACACGGTGGGCGGAGCGGCC	14
AG31 16992_1	TGGTTTGGCTTTTACCAAACTGGTGGTGTAGGGGTTTCGGGTTCTGTAGCTACGTACTCT	13
BS19 9145_2	TGGCTAGGGTTCTATATGAGCGGTGGTATAGGGTTCTCAAACACAGCGGGCGGAGCGGCC	13
BB3 1324_1	TGGCTGGGGTTCTACATGAGCGGCGGCATA----------	NA
BB3 5336_3	TGGCTTGGATTCTACATGAGCGGCGGCATAGGTTTCTCAAATACCGTCGCCGCAGCAGCT	13
BV9 1718_1	TGGCTAGGGTTCTATATGAGCGGTGGTATAGGGTTCTCAAACACAGTGGGCGGAGCGGCC	14
BV3 6795_1	----------	
CX10 14344_1	TGGCTTGGATTTTACATGAGTGGCGGTATCGGTTTCTCAAACACAACTACTGCAGCCGCT	11
CX10 1922_1	TGGCTTGGATTCTACATGAGCGGCGGCATAGGTTTTTCAAATACCGTCGGCGCCTCGGGT	15
COP3-100133354	TGGTTGGGGTTCTACATGAGTGGGGGGATTGGATTCTCAAATACCGTAGCTACATCTGCT	13
COP3-100174111	TGGTATGGCCAGTACATGGCAGGAGGTATTGGCTTTAATGCGGCAGGATACTCAATACTT	17
COP3-100202833	TGGTACGGTCAGTACATGGCAGGTGGTGTCGGATTCAATGCAGCCGCATACTCTTTCCTC	10
COP3-100352853	TGGTTAGGACAGTTTATGGCCGGAGGTATTGGCTTTCCCTCTGCAGGACTTGGACTTCTT	18
COP3-100679051	TGGTATGGTCAACAGATGGCCGGAGGAATCGGATTCCCATCTGCAGGATATGCTTTCCTT	14

	MLr AACTAYCCSAACTAYGCAATGAA (reverse complemented)	Mismatches
M_smithii	AGAGGACCAAACTATCCGAACTATGCTATGAATGTAGGTCACCAGCCAGAATATGCAGGT	1
M_paldcla	AGAGGCCCGAACTACCCGAACTACGCGATGAACGTCGGCCACCAGGGTGGCTACGCCGGT	1
M_cncili	AGAGGCGCCAACTACCCGAACTACGCCATGAACGTGGGCCACCAGGGCGAGTACGCAGGC	1
M_hngti	CGTGGTCCAAACTACCCGAACTACGCAATGAACGTGGGTCACCAGGGAGAATACGCAGCA	0
M_holnda	CGTGGACCAAACTACCCGAACTACGCCATGAACGTAGGTCACCAGGGTGGATACGCTGCA	1
M_jannshi	AGAGGGCCTAACTATCCAAACTACGCAATGAACGTTGGTCACCAAGGAGAATATGCTGGA	1
M_instststns	AGAGGACCCAACTATCCAAACTATGCAATGAACGTTGGTCACCTTTGCGGTTACGCTGGT	1
M_thrmoauto	AGAGGAGCTAACTACCCCAACTACGCCATGAATGTGGGTCACCAGGGTGAATATGCAGGT	1
M_stdtm	AGAGGTCCTAACTACCCTAACTACGCAATGAACGTAGGTCACCAACCTGAATATGCAGGT	1
M_barkeri	CGTGGTCCAAACTACCCCAACTACGCAATGAACGTTGGTCACCAGGGTGGATACGCAGGT	0
M_kandleri	CGTGGTCCGAACTACCCGAACTACGCGATGAACGTGGGTCACCTGGGCGAGTACGCGGGT	1
BA1 24_9	AGGGGCTACAACACCCCATACGCATCTTTCACGGCTGGGCACGGCGTCATCCGGGAGGTG	12
BA2 13_151	AGAGGTCTTAACTACCCGATGCAATCATTCAGCGCAGCCCATGGTTCTCTAAGGGATGCA	9
BV3 3440_1	AGAGGCCTTAACTACCCGATGCAAACTTTCACAGCAGCCCACGGTGTCCTC---------	9
CD8 186481_1	AGAGGTCTTAACTATCCGATGCAATCTTTCAGTGCAGCGCACGGTGCTATAAGGGATACG	9
BS36 5099_1	----------	
BS36 20936_1	AGAGGTCTTAACTATCCGATGCAATCTTTTAGTGCAGCGCACGGTGCTATAAGGGATACG	9
AG31 16992_1	CGTGGGATGAACTATCCAATGCACGCCTTCACGCCAGCGCACGCAGCTATGAGGCAAGCT	8
BS19 9145_2	AGAGGTCTTAACTATCCGATGCAATCTTTTAGTGCAGCGCACGGTGCTATAAGGGATACG	9
BB3 1324_1	----------	
BB3 5336_3	AGAGGCCTTAACTACCCGATGCAAACTTTCACGGCAGCCCACGGTGTCCTCCGAGATACG	9
BV9 1718_1	AGAGGTCTTAACTATCCGATGCAATCTTTTAGTGCAGCGCACGGTGCTATAAGGGATACG	9
BV3 6795_1	----------	
CX10 14344_1	AGGGGTCTAAATTACCCAATGCAATCGTTCAGCGCGGCTCATGGTGCTCTTAGGGCTGCG	11
CX10 1922_1	----------	
COP3-100133354	----------	
COP3-100174111	AAAGGTCCGAATGTTCCATACATGTCATATACAGTCGGTGCACTTGGGTTCCCGATGTTC	13
COP3-100202833	AAAGGTCCGAATGTACCATACATGTCGTACACTGTTGGCGCTCTTGGGATGCCGATGTTC	15
COP3-100352853	AAAGGAGAGAATGTGCCATACATGTCCTATACCGCAGTAGGTGGTAGCCTTGGTGTAGGA	15
COP3-100679051	----------	

Fig. S5.

Binding sites of commonly used *mcrA* gene primers. Sites for the primer pairs of (A) Hales *et al*. (*76*), (B) Springer *et al*. (*77*) and (C) Luton *et al*. (*78*) against representative euryarchaeotal methanogens and bathyarchaeotal *mcrA* sequences. Bases colored green match primer sequences, bases colored red do not match primer sequences, and bases colored brown are primer sequences. M_smithii, *Methanobrevibacter smithii* PS; M_paldcla, *Methanocella paludicola* SANAE; M_cncili, *Methanosaeta concilii* GP6; M_hngti, *Methanosprilium hungatei* JF-1; M_holnda, *Methanomethylovorans* hollandica DMS1; M_jannshi, *Methanocaldococcus jannaschii* JAL-1; M_instststns, *Candidatus* Methanomassiliicoccus intestinalis Issoire-Mx1; M_thrmoauto, *Methanothermobacter thermautotrophicus* ΔH; M_stdtm, *Methanosphaera stadtmanae* MCB-3; M_barkeri, *Methanosarcina barkeri* FUSARO; M_kandleri, *Methanopyrus kandleri* AV19.

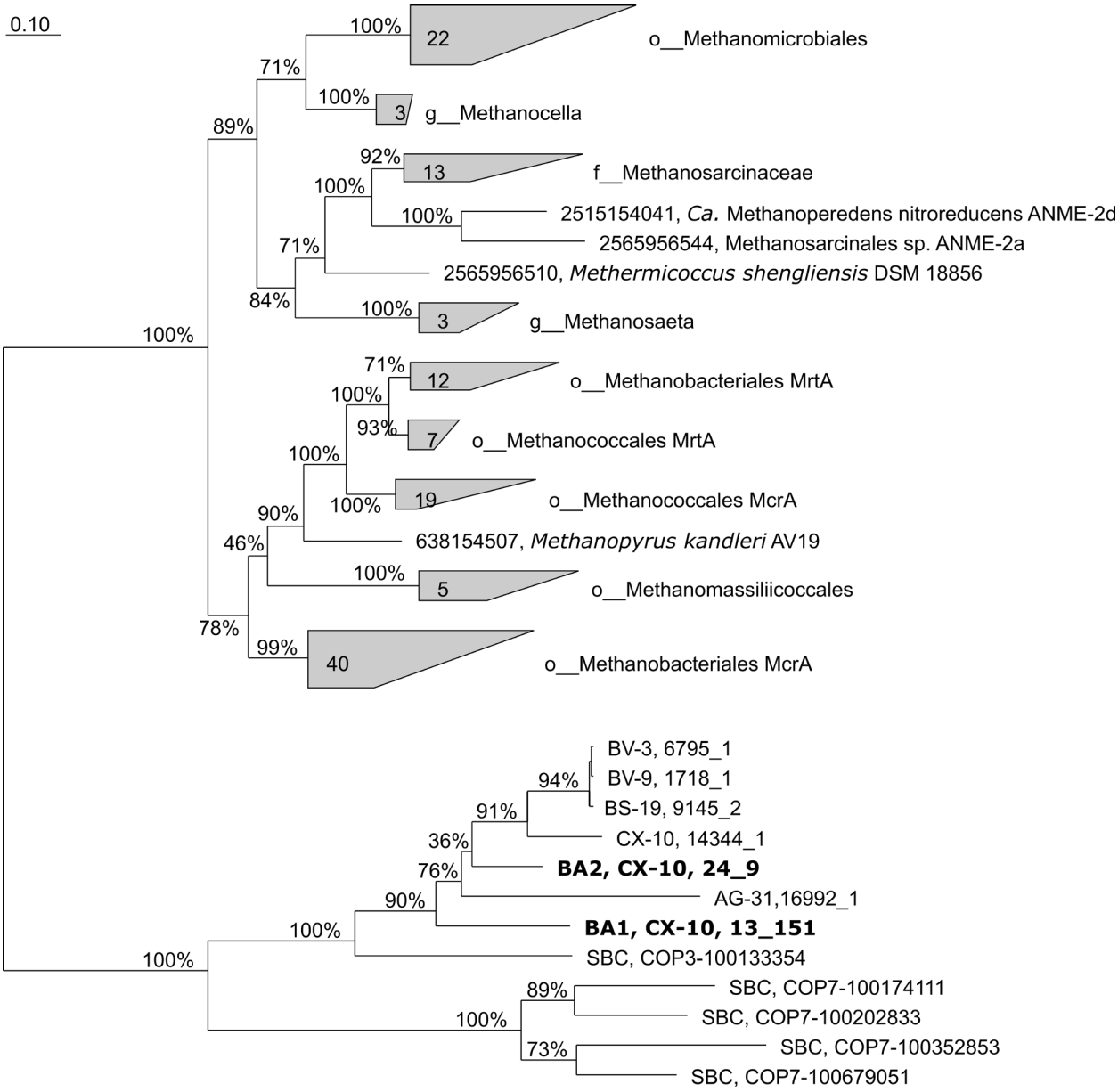


Fig. S6.

Maximum likelihood tree for methyl-coenzyme M reductase subunit A (McrA) amino acid sequences identified within the Surat Basin, Coal Oil Point, and IMG. Inference was performed with FastTree under the WAG+G model from a multiple sequence alignment obtained using MAFFT. Sequences were grouped at the order, family, or genus level in order to show the relationship between major clades. The number of taxa within each collapsed group is indicated within the quadrilaterals. Stability of the tree topology was evaluated using 100 non-parameteric bootstrap replicates and is shown along each branch. Sequences within the BA1 and BA2 genomes are shown in bold.

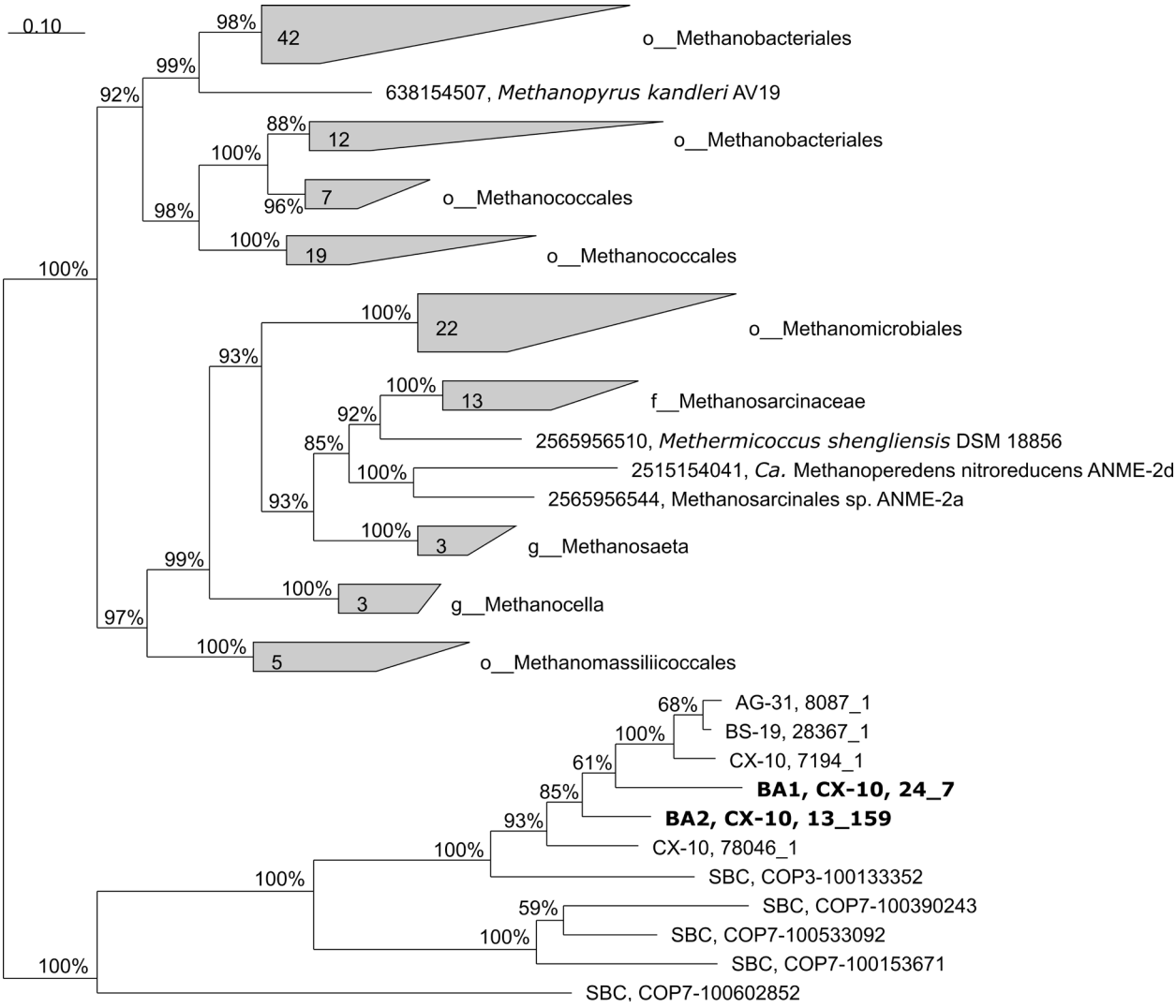


Fig. S7.

Maximum likelihood tree for methyl-coenzyme M reductase subunit B (McrB) amino acid sequences identified within the Surat Basin, Coal Oil Point, and IMG. Inference was performed with FastTree under the WAG+G model from a multiple sequence alignment obtained using MAFFT. Sequences were grouped at the order, family, or genus level in order to show the relationship between major clades. The number of taxa within each collapsed group is indicated within the quadrilaterals. Stability of the tree topology was evaluated using 100 non-parameteric bootstrap replicates and is shown along each branch. Sequences within the BA1 and BA2 genomes are shown in bold.

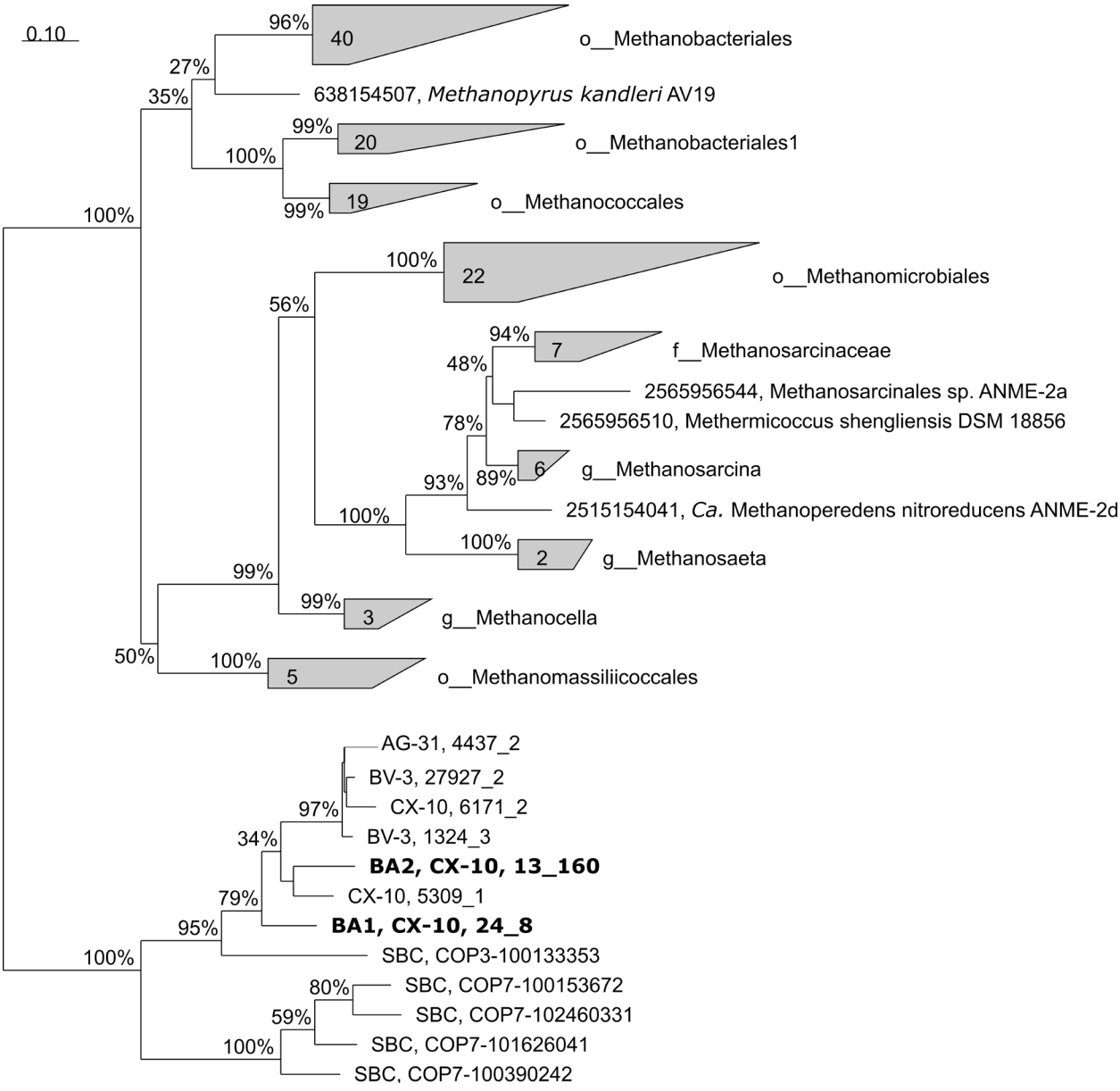


Fig. S8.

Maximum likelihood tree for methyl-coenzyme M reductase subunit gamma (McrG) amino acid sequences identified within the Surat Basin, Coal Oil Point, and IMG. Inference was performed with FastTree under the WAG+G model from a multiple sequence alignment obtained using MAFFT. Sequences were grouped at the order, family, or genus level in order to show the relationship between major clades. The number of taxa within each collapsed group is indicated within the quadrilaterals. Stability of the tree topology was evaluated using 100 non-parameteric bootstrap replicates and is shown along each branch. Sequences within the BA1 and BA2 genomes are shown in bold.

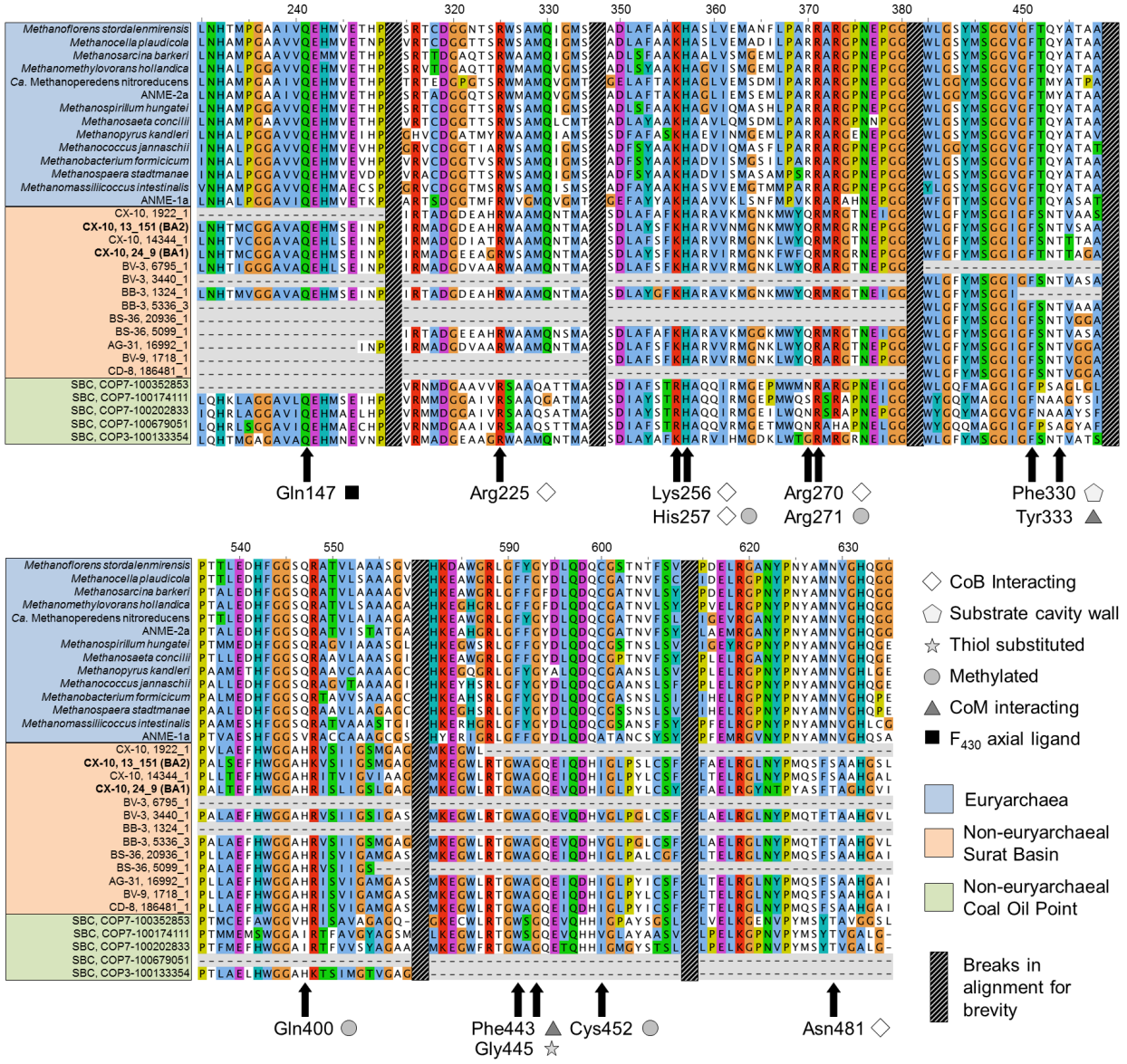


Fig. S9.

Conservation of functionally important McrA residues. Ermler *et al*. (*22*) identified important residues based on their position in and around the Mcr active site as well as their interaction with the three main cofactors: CoM, CoB and F_{430}. The residue numbering system follows Ermler *et al*. (*22*).

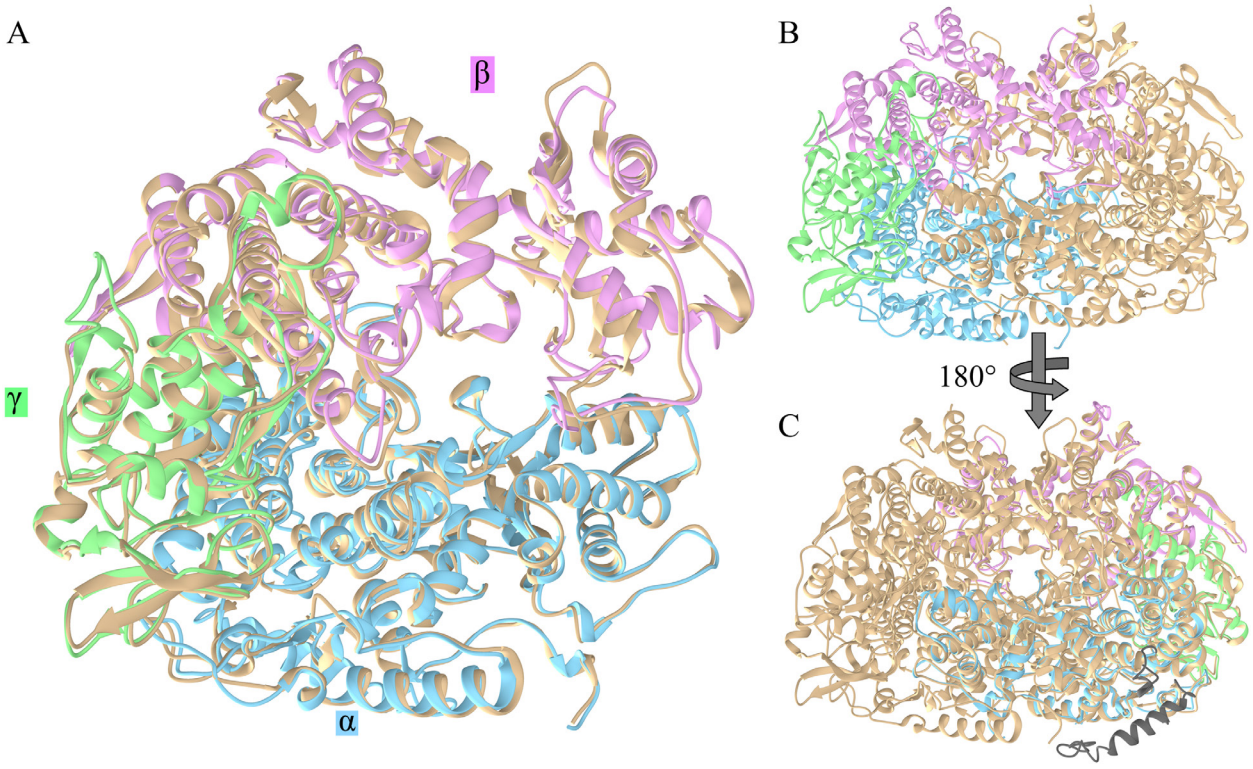


Fig. S10.

Structural model of the BA1 Mcr complex. (A) Structural models of the α (blue), β (purple) and γ (green) subunits superimposed on the corresponding subunits from the *Methanopyrus kandleri* crystal structure (PDB ID: 1E6V, shown in tan). (B) Dimeric $\alpha_2\beta_2\gamma_2$ Mcr complex with the α, β and γ subunits from the BA1 model and the α', β' and γ' subunits from the *Methanopyrus kandleri* crystal structure reveals high structural similarity between the model and crystal structure. (C) The N-terminal extension and insertion discussed in the main text are highlighted in grey. The structural model suggests that these extra amino acids, while separated by ~300 amino acids in the primary sequence, are quite close to one another on the surface of the Mcr complex, suggesting possibly interactions with other proteins in the genome. These extra amino acids do not interfere with the active site or interactions between the α subunit and the rest of the complex.

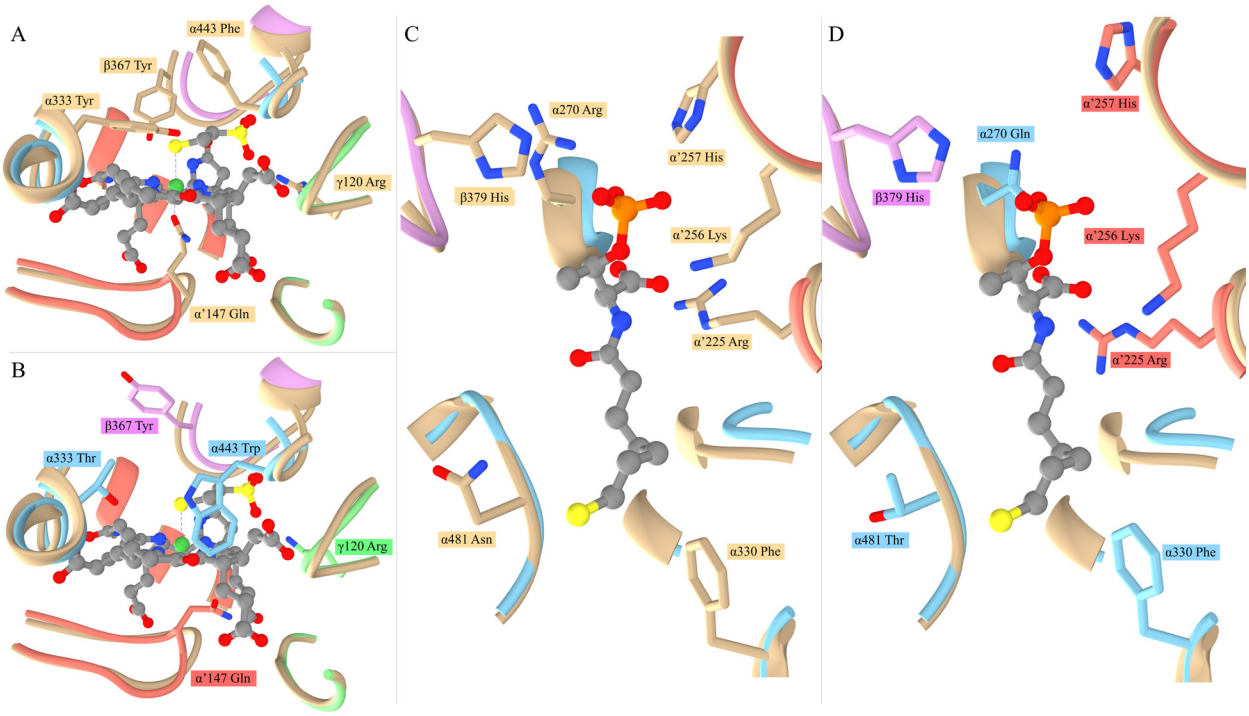


Fig. S11.

Conserved active site residues in the BA1 Mcr model. Mcr α, β and γ subunits from the BA1 genome and the *Methanopyrus kandleri* crystal structure are shown using the same color scheme as in **Fig. S10**. Red, yellow, and orange spheres represent the oxygen atom of –OH hydroxyl groups, sulfur, and phosphorus, respectively. The *M. kandleri* and BA1 Mcr backbones in close proximity to the CoM, CoB and F_{430} cofactors are shown as ribbons. (A,B) The side chains of residues interacting with the CoM and F_{430} cofactors from Ermler *et al*. (*22*) are shown for the *M. kandleri* and BA1 Mcrs, respectively. (C,D) The side chains of residues interacting with the CoB cofactor from are shown for the *M. kandleri* and BA1 Mcrs, respectively. The majority of amino acid residues important for cofactor binding are identical between the BA1 and euryarchaeotal Mcrs, including four of the five residues making salt bridges with the CoB phosphate group and the axial ligand for the F_{430} Ni atom (note that the orientation of the side chains of the structural model should not be considered as I-TASSER does not optimize these based on interactions with cofactors). Noteworthy substitutions between the euryarchaeotal and BA1 Mcrs occur at positions α443, α270, α481 and α333, but for the most part the substituted residues can carry out functions similar to those predicted for the euryarchaeotal sequences. The substitution of Phe with Trp at position α443 maintains the bulky aromatic nature of the residue, which is thought to be the main function of α443, pinning CoM between the side chain aromatic and the F_{430}. The substitution of Tyr with Thr at position α333 maintains the hydroxyl functional group, albeit with smaller side chain volume, and without the potential for deprotonation which comes with the Tyr hydroxyl group being attached to an aromatic ring. Whether the ability of the α333 Tyr to be deprotonated is important for Mcr function is unknown. The substitution of Asn with Thr at position α481 replaces one non-charged polar residue with another. It is possible the polar hydroxyl group of Thr could interact with the thiol of CoB in a similar way as the polar amine group of Asn. The substitution of Arg with Gln at position α270 removes the ability of the

residue at this position to form a salt bridge with CoB's phosphate group because Gln is not charged. This is only one of the five residues which forms a salt bridge with CoB in euryarchaeotal Mcr, and the amine group of Gln could still donate a hydrogen bond to a phosphate oxygen of CoB.

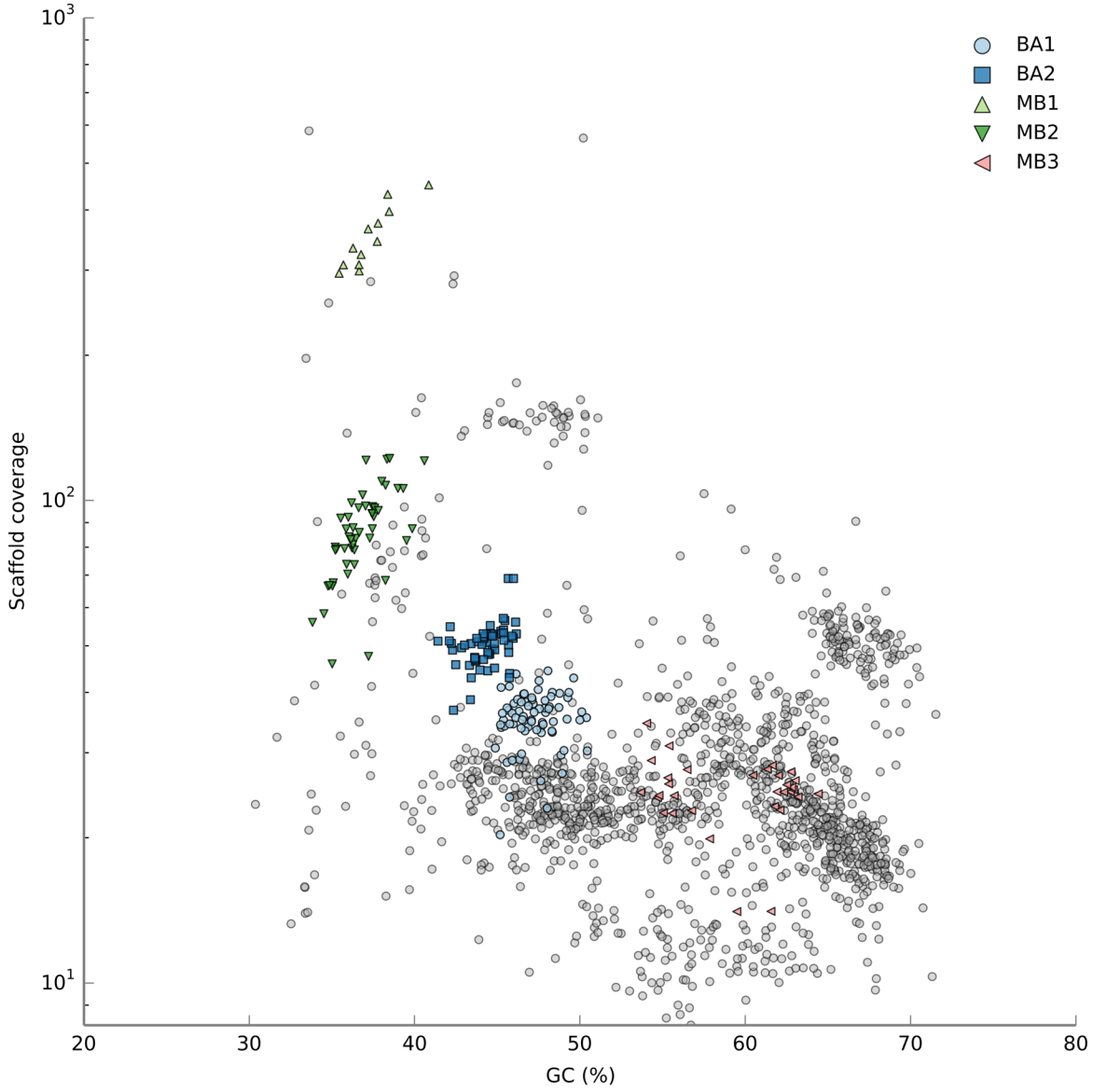


Fig. S12.

Percentage GC and coverage of scaffolds from *Bathyarchaeota* and *Euryarchaeota* genomes recovered from the CX-10 coalbed methane well. Each point represents a scaffold plotted as a function of its average GC-content (x-axis) and coverage (y-axis). Scaffolds from the bathyarchaeotal (BA1 and BA2) and 3 euryarchaeotal genomes (MB1, MB2, and MB3) within CX-10 are highlighted with different colors, and the remaining scaffolds shown in grey. For clarity, only scaffolds > 10kb are shown. The wide GC-content distribution of the MB3 genome suggests it may be chimeric, but it is shown to illustrate that all euryarchaeotal genomes are well-separated from BA1 and BA2.

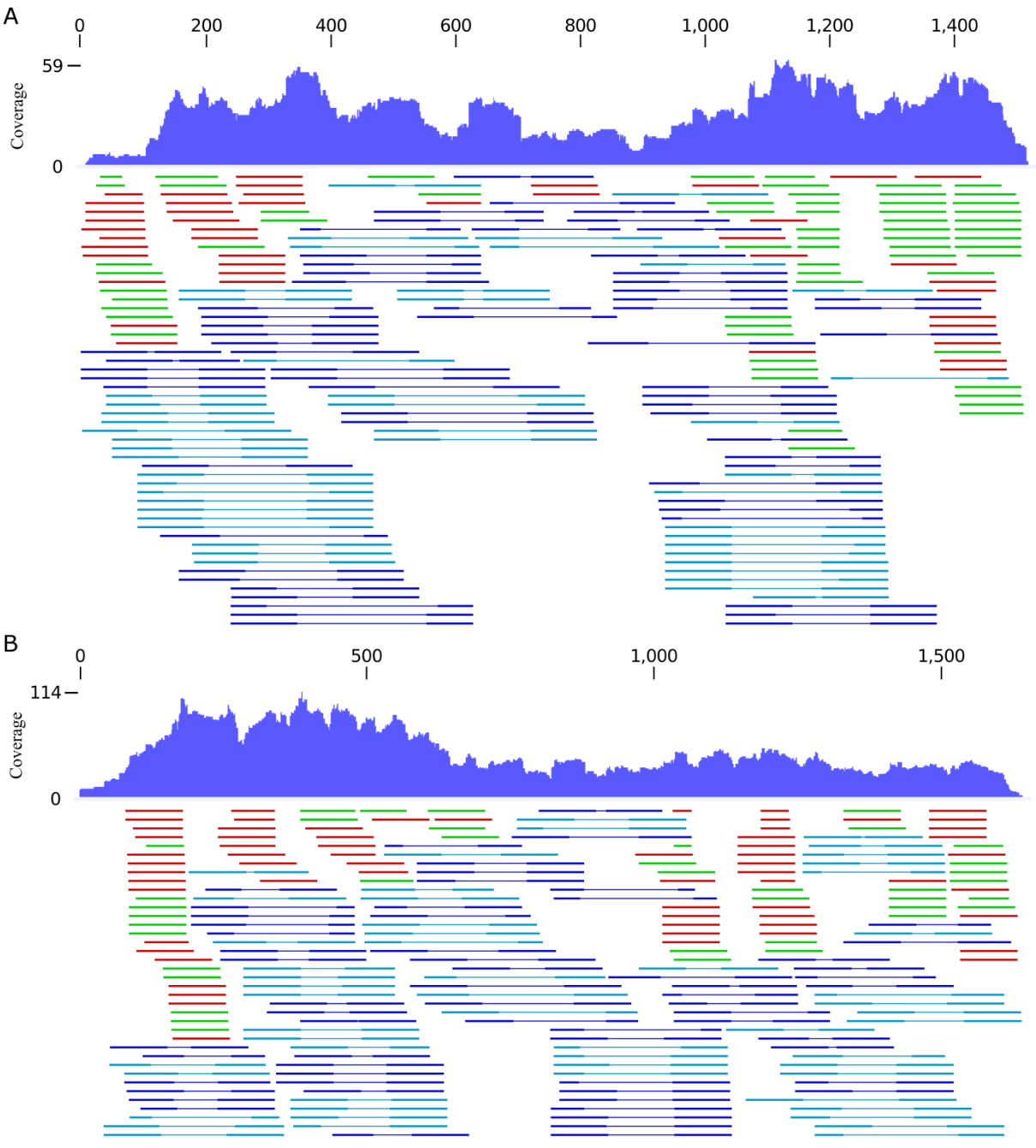


Fig. S13.

Reads mapping to the 16S rRNA gene from (A) BA1 and (B) BA2. Reads in blue illustrate paired reads mapped in the forward (dark blue) or reverse (light blue) orientation. Reads in green and red are single reads mapped in the forward and reverse orientation, respectively. Only a small subset of mapped reads are shown, but these are sufficient to illustrate that paired reads span the entire length of both 16S rRNA genes.

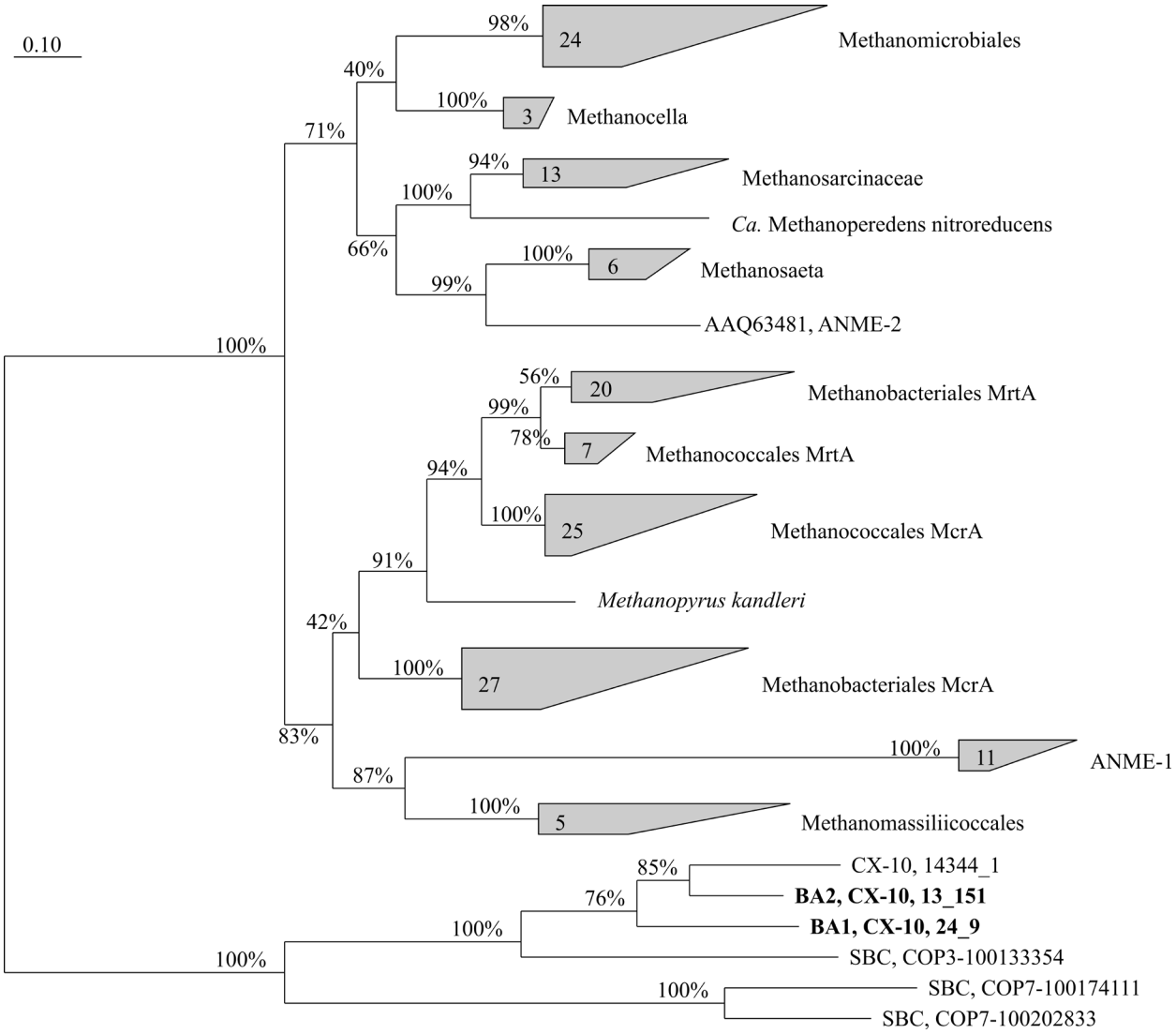


Fig. S14.

Maximum likelihood tree for methyl-coenzyme M reductase subunit A (McrA) proteins. The tree was inferred across near full length McrA proteins identified within the Surat Basin and Coal Oil Point along with 153 additional proteins obtained from GenBank. All sequences are ≥400 amino acids in length. Inference was performed with RAxML under the LG+I+F+G model from a multiple sequence alignment obtained using FSA. Sequences were grouped at the order or genus level in order to show the relationship between major clades including known archaeal methanotrophs. The number of taxa within each collapsed group is indicated within the quadrilaterals. Stability of the tree topology was evaluated using 100 non-parametric bootstrap replicates and is shown along each branch. Sequences within the BA1 and BA2 genomes are shown in bold.

Table S1.

Characteristics of population genomes recovered from the CX-10 coalbed methane well.

Bin Id	Putative taxonomy	Comp.	Cont.	Genome size (bp)	no. scaffolds	no. contigs	%GC	no. CDS	Coverage
BA2	p__Bathyarchaeotal	93.8	3.7	1,455,689	57	58	44.2	1761	49.8
MB1	f__Methanobacteriaceae	92.0	1.6	2,128,045	12	12	37.6	2047	360.6
BA1	p__Bathyarchaeotal	91.6	2.8	1,931,714	89	96	47.1	2403	35.8
MB2	f__Methanobacteriaceae	86.1	0.0	1,353,497	90	93	36.8	1448	85.3
BC1	k__Bacteria	84.8	0.0	1,816,487	72	74	65.4	1808	51.7
DP1	c__Deltaproteobacteria	80.2	3.2	2,505,846	80	99	50.2	2429	22.1
SP1	c__Spirochaetia	79.1	0.0	972,899	66	73	47.7	1025	24.0
AB1	c__Actinobacteria	75.0	5.7	1,957,451	55	63	63.8	1895	23.4
MP1	o__Methylophilales	63.8	0.0	1,354,941	64	69	55.0	1416	23.5
BC2	k__Bacteria	62.1	0.0	447,382	19	19	38.9	457	78.4
MC1	f__Methylococcaceae	50.0	0.0	1,328,640	79	83	49.2	1191	25.3
DP2	c__Deltaproteobacteria	47.7	3.6	1,619,427	24	25	48.8	1554	147.9
DP3	c__Deltaproteobacteria	46.6	0.0	713,283	14	14	45.2	702	143.5
MB3	f__Methanobacteriaceae	44.6	2.8	696,060	55	62	59.6	749	24.7
BC3	k__Bacteria	37.4	1.7	2,144,441	101	143	65.9	1862	18.9
AB2	c__Actinobacteria	30.3	0.0	866,671	52	53	68.3	870	51.6
RC1	f__Rhodocyclaceae	27.6	3.5	1,116,001	72	76	64.8	1168	24.7
PB1	p__Proteobacteria	26.7	2.1	1,805,719	211	276	66.9	1962	15.4
RC2	f__Rhodocyclaceae	26.1	0.0	883,429	57	59	62.3	927	38.1

Table S2.

Amino acid and 16S rRNA identity of representatives from the archaeal phylum *Bathyarchaeota*.

AAI \ shared CDS	BA1	BA2	E09
BA1	-	656	97
BA2	67.0%	-	88
E09	45.7%	47.4%	-

16S rRNA identity \ alignment length	BA1	BA2	E09
BA1	-	1382	899
BA2	90.5%	-	762
E09	81.5%	83.5%	-

Table S3.

Genes shared by BA1, BA2, and E09 genomes.

BA1 gene id	BA2 gene id	E09 gene id	Annotation
ba1_4_9	ALXK01000078.1_52	ba2_14_2	30S ribosomal protein S17P
ba1_6_30	ALXK01000031.1_8	ba2_42_3	30S ribosomal protein S19
ba1_51_29	ALXK01000078.1_23	ba2_40_20	30S ribosomal protein S24e
ba1_4_13	ALXK01000078.1_49	ba2_46_1	30S ribosomal protein S4e
ba1_4_20	ALXK01000078.1_41	ba2_46_9	30S ribosomal protein S5
ba1_4_15	ALXK01000078.1_46	ba2_46_4	30S ribosomal protein S8
ba1_83_12	ALXK01000025.1_5	ba2_6_9	30S ribosomal protein S8e
ba1_4_10	ALXK01000078.1_51	ba2_14_1	50S ribosomal protein L14
ba1_4_25	ALXK01000078.1_39	ba2_46_12	50S ribosomal protein L15P
ba1_4_19	ALXK01000078.1_42	ba2_46_8	50S ribosomal protein L18P
ba1_4_18	ALXK01000078.1_43	ba2_46_7	50S ribosomal protein L19e
ba1_6_29	ALXK01000031.1_7	ba2_42_4	50S ribosomal protein L2
ba1_4_4	ALXK01000031.1_9	ba2_14_6	50S ribosomal protein L22P
ba1_6_28	ALXK01000031.1_6	ba2_42_6	50S ribosomal protein L23P
ba1_4_6	ALXK01000078.1_54	ba2_14_4	50S ribosomal protein L29P
ba1_6_26	ALXK01000031.1_4	ba2_42_8	50S ribosomal protein L3
ba1_4_21	ALXK01000078.1_40	ba2_46_10	50S ribosomal protein L30P
ba1_4_17	ALXK01000078.1_44	ba2_46_6	50S ribosomal protein L32e
ba1_6_27	ALXK01000031.1_5	ba2_42_7	50S ribosomal protein L4
ba1_4_16	ALXK01000078.1_45	ba2_46_5	50S ribosomal protein L6
ba1_22_2	ALXK01000002.1_1	ba2_2_19	CCA-adding enzyme
ba1_65_45	ALXK01000025.1_8	ba2_6_12	Chromosome partition protein Smc
ba1_51_24	ALXK01000078.1_20	ba2_40_17	DNA-directed RNA polymerase subunit E'
ba1_11_30	ALXK01000074.1_5	ba2_4_32	Endonuclease NucS
ba1_46_8	ALXK01000043.1_4	ba2_19_28	FUN14 family protein
ba1_16_45	ALXK01000031.1_3	ba2_54_7	G/T mismatches repair enzyme
ba1_83_10	ALXK01000025.1_3	ba2_6_7	H/ACA RNA-protein complex component Gar1
ba1_13_54	ALXK01000078.1_34	ba2_5_27	Histidine--tRNA ligase
ba1_14_17	ALXK01000002.1_2	ba2_8_57	hypothetical protein
ba1_33_10	ALXK01000025.1_10	ba2_9_21	hypothetical protein
ba1_26_9	ALXK01000061.1_9	ba2_4_31	hypothetical protein
ba1_14_43	ALXK01000074.1_18	ba2_8_39	hypothetical protein
ba1_40_16	ALXK01000074.1_50	ba2_17_75	hypothetical protein
ba1_51_26	ALXK01000078.1_22	ba2_40_19	hypothetical protein
ba1_2_40	ALXK01000011.1_9	ba2_47_21	molybdenum cofactor biosynthesis protein A
ba1_4_26	ALXK01000078.1_38	ba2_46_13	preprotein translocase subunit SecY
ba1_2_43	ALXK01000045.1_10	ba2_22_17	putative phosphoglucosamine mutase
ba1_87_36	ALXK01000022.1_10	ba2_3_8	Pyruvate kinase, alpha/beta domain
ba1_2_33	ALXK01000077.1_3	ba2_51_22	Pyruvate synthase subunit PorD
ba1_4_8	ALXK01000078.1_53	ba2_14_3	Ribonuclease P protein component 1
ba1_83_14	ALXK01000025.1_6	ba2_6_10	segregation and condensation protein B
ba1_83_11	ALXK01000025.1_4	ba2_6_8	Signal recognition particle 19 kDa protein
ba1_13_21	ALXK01000061.1_7	ba2_12_50	single-stranded DNA-binding protein
ba1_77_14	ALXK01000074.1_19	ba2_13_13	TATA-box-binding protein
ba1_13_22	ALXK01000061.1_6	ba2_12_49	Threonine synthase
ba1_51_25	ALXK01000078.1_21	ba2_40_18	Transcription elongation factor Spt4
ba1_16_11	ALXK01000045.1_6	ba2_1_12	Transcription factor E
ba1_83_9	ALXK01000025.1_2	ba2_6_6	Transcription initiation factor IIB
ba1_34_2	ALXK01000026.1_3	ba2_15_30	Translation initiation factor 1A
ba1_51_38	ALXK01000045.1_9	ba2_19_36	Triosephosphate isomerase
ba1_33_12	ALXK01000025.1_9	ba2_9_19	Tritrans,polycis-undecaprenyl-diphosphate synthase
ba1_16_13	ALXK01000045.1_7	ba2_1_11	tRNA (cytidine(56)-2'-O)-methyltransferase
ba1_16_10	ALXK01000045.1_5	ba2_1_13	tRNA 2'-O-methylase
ba1_13_55	ALXK01000078.1_33	ba2_5_28	tRNA-splicing endonuclease
ba1_66_2	ALXK01000047.1_1	ba2_15_26	Type 2 DNA topoisomerase 6 subunit A
ba1_15_2	ALXK01000076.1_6	ba2_47_11	VCP-like ATPase

Table S4.

Scaffolds in the BA1 and BA2 genomes containing one or more genes associated with archaeal methane metabolism.

Gene	BA1 Genes	BA1 Scaffold(s)	BA2 Genes	BA2 Scaffold(s)
Formylmethanofuran dehydrogenase (Mo-containing) (*fwd*)	*fwdABCD*	89	-	-
Formylmethanofuran dehydrogenase (W-containing) (*fmd)*	-	-	-	-
Formylmethanofuran:H_4MPT formyltransferase (*ftr*)	*ftr*	58, 89	*ftr*	10
Methenyl-H_4MPT cyclohydrolase (*mch*)	*mch*	89	-	-
Methylene- H_4MPT dehydrogenase (*mtd*)	*mtd*	89	-	-
F_{420}-reducing hydrogenase (*frh*)	*frhABG*	89, 70	-	-
H_2-forming methylene dehydrogenase (*hmd*)	-	-	-	-
Methylene- H_4MPT reductase (*mer*)	*mer*	25, 44, 81	-	-
Methyl- H_4MPT: coenzyme M methyltransferase (*mtr*)	*mtrH*	15, 45, 89	-	-
Methyl-coenzyme M reductase (*mcr*)	*mcrABG*	24	*mcrABCDG*	13
Methyl-coenzyme M reductase II (*mrt*)	-	-	-	-
Heterodisulfide reductase (cytoplasmic) (*hdr*)	*hdrABC*	10, 52, 87	*hdrABC*	10, 13, 15, 22
Heterodisulfide reductase (membrane bound) (*hdr*)	*hdrD*	3, 66, 70, 87	*hdrD*	13, 22, 43
Non-F_{420}-reducing hydrogenase (soluble, minus cytochrome) (*mvh*)	*mvhADG*	66, 70	*mvhADG*	22
Non-F_{420}-reducing hydrogenase (soluble, minus cytochrome) (*mvh*)	*mvhB*	85, 87	*mvhB*	8, 22
Non-F_{420}-reducing hydrogenase (membrane bound, cytochrome containing) (*vht*, *vho*)	-	-	-	-
Energy converting hydrogenase (*ech*)	*echABCDEF*	2	-	-
Energy converting hydrogenase (*eha*, *ehb*)	-	-	-	-
Methanol methyltransfersase (*mta*)	*mtaA*	45	-	-
Methylamine methyltransferase (*mtb*)	*mtbA*	45	-	-
Methylsulfide methyltransferase (*mts*)	*mtsA*	45	-	-
Trimethylamine corrinoid (*mtt*)	*mttBC*	45, 89	-	-
Dimethylamine corrinoid (*mtb*)	*mtbBC*	5, 15, 45, 87	-	-
Carbon monoxide dehydrogenase (*coo*)	*CooSC*	70	-	-
Carbon monoxide dehydrogenase (*cdh*)	*cdhABDGE*	5	*cdhABDGE*	55
Acetyl-CoA synthase (ADP-forming) (*acd*)	*acd*	57, 67, 87	*acd*	19
Acetyl-CoA synthase (*acs*)	-	-	*acs*	4
Pyruvate Fd:oxidoreductase (*por*)	*porABDG*	2	*porABDG*	51
Ketoglutarate Fd:oxidoreductase (*kor*)	*korABC*	42	-	-
Indolepyruvate Fd:oxidoreductase (*ior*)	*iorAB*	50	-	-
Valerate Fd:oxidoreductase (*vor*)	*vorABCG*	46	-	-

Table S5.

Characteristics of non-euryarchaeotal *mcrA* genes identified in the Surat Basin and Coal Oil Point metagenomes.

Sample site	Gene identifier	Scaffold length (bp)	Gene length (bp)	Coverage	Note
AG-31	16992_1	922	921	4.2	partial
BB-3	1324_1	3,418	813	11.4	partial
BB-3	5336_3	2,854	840	13.2	partial
BV-3	6795_1	1,052	987	14.7	partial
BV-3	3440_1	560	558	3.4	partial
BV-9	1718_1	2,230	996	13.9	partial
BS-19	9145_2	1,807	1248	5.2	partial
BS-36	20936_1	1,505	798	3.5	partial
BS-36	5099_1	1,135	693	3.5	partial
CX-10	24_9	9,346	1815	41.9	complete; contained in BA1
CX-10	13_151	154,546	1824	50.4	complete; contained in BA2
CX-10	14344_1	4,802	1824	21.9	complete
CX-10	1922_1	818	816	91.4	partial
CD-8	186481_1	721	720	3.5	partial
COP-3	JGI24025J20009_100133354	4,354	1488	25.5	partial
COP-7	JGI24724J26744_100174111	3,721	1869	22.5	complete
COP-7	JGI24724J26744_100202833	3,352	1932	25.8	complete
COP-7	JGI24724J26744_100352853	2,273	1209	19.9	partial
COP-7	JGI24724J26744_100679051	1,397	1347	19.1	partial

Table S6.

Reads with highest sequence similarity to non-euryarchaeotal Surat Basin or Coal Oil Point *McrA* proteins.

SRA Run	Fragment Id	Best hit	% identity	Alignment length (aa)	Subject start	Subject end	e-value	Bit score
Sediment sampled at 10cm bsf from a freshwater lake in Sweden (Illumina HiSeq 2000)								
ERR476714	23183389	BB-3_5336_3	66.7	33	248	280	6.0E-09	52.4
ERR476714	62450462	BV-3_6795_1	71.8	39	131	169	3.0E-11	60.1
ERR476714	4946464	BV-3_6795_1	66.7	33	154	186	1.0E-08	51.2
ERR476714	62050423	CX-10_14344_1	67.7	34	236	269	2.0E-09	54.3
ERR476714	52841199	BA1_24_9 (CX-10)	71.0	31	423	453	1.0E-09	55.1
ERR476714	32445081	JGI24025J20009_100133354	71.9	32	280	311	1.0E-08	51.2
ERR476714	61663865	JGI24724J26744_100202833	68.3	41	156	196	1.0E-11	61.2
ERR476714	2549976	JGI24724J26744_100202833	75.0	32	164	195	1.0E-10	58.2
ERR476714	39339866	JGI24724J26744_100202833	52.6	38	169	206	6.0E-09	52.4
Sediment sampled at 0-4 cm or 10-15 cm bsf from the Tonya Seep, Santa Barbara Channel (454 GS FLX Titanium)								
SRR099552	102336	JGI24724J26744_100679051	60.8	74	3	72	7.0E-21	93.2
SRR099554	50075	JGI24724J26744_100174111	83.1	83	427	509	2.0E-38	151.4
Sediment from a marine fish farm (Illumina HiSeq 2000)								
SRR1022349	2008616	CX-10_14344_1	79.4	34	568	601	7.0E-10	55.5
SRR1022349	4136140	BA2_13_151 (CX-10)	75.8	33	468	500	2.0E-10	57.8
SRR1022349	9684352	JGI24724J26744_100174111	68.8	32	298	329	3.0E-08	50.1
SRR1022349	2909930	JGI24724J26744_100202833	76.7	30	584	613	2.0E-09	54.3
SRR1022349	1173094	JGI24724J26744_100352853	70.2	57	275	327	7.0E-14	68.9
SRR1022349	10162529	JGI24724J26744_100352853	79.4	34	361	394	3.0E-10	56.6
SRR1022349	8440102	JGI24724J26744_100352853	68.8	32	95	126	5.0E-09	52.8
Suncor tailing pond in North Alberta, Canada sampled at 5 mbs (Illumina HiSeq 2000)								
SRR1313281	18090878	JGI24025J20009_100133354	83.7	49	386	434	2.0E-19	87.8
SRR1313281	1951519	JGI24025J20009_100133354	77.1	48	326	373	1.0E-17	81.6
SRR1313282	23431410	JGI24025J20009_100133354	83.9	62	112	170	6.0E-24	102.8
SRR1313282	7412287	JGI24025J20009_100133354	77.6	49	318	366	8.0E-19	85.9
SRR1313283	30372638	BB-3_5336_3	64.0	50	229	278	4.0E-15	72.4
SRR1313283	3553601	BA2_13_151 (CX-10)	69.4	49	439	487	3.0E-18	82.8
SRR1313283	40152408	BA2_13_151 (CX-10)	73.5	49	552	600	2.0E-16	77.0
SRR1313285	27787210	CX-10_14344_1	67.4	49	545	593	4.0E-15	73.6
SRR1313285	40936392	JGI24025J20009_100133354	98.0	49	175	223	8.0E-24	102.4
SRR1313285	30980861	JGI24025J20009_100133354	73.6	53	51	103	4.0E-20	90.1
SRR1313285	36554350	JGI24025J20009_100133354	66.2	65	434	493	9.0E-20	89.0
SRR1313285	35468171	JGI24025J20009_100133354	54.4	46	258	303	9.0E-09	52.4
SRR1313287	11167557	JGI24025J20009_100133354	92.0	50	234	283	7.0E-23	99.4
SRR1313287	7263386	JGI24025J20009_100133354	91.8	49	267	315	4.0E-22	96.7
SRR1313287	39143853	JGI24025J20009_100133354	78.0	50	67	116	5.0E-19	86.7
Suncor tailing pond in North Alberta, Canada sampled at 6 mbs (Illumina HiSeq 2000)								
SRR942941	4298686	JGI24025J20009_100133354	94.0	50	269	318	4.0E-23	100.1
SRR943329	42739088	JGI24025J20009_100133354	80.0	50	44	93	3.0E-21	90.5
SRR943329	3732144	JGI24025J20009_100133354	85.1	47	35	81	3.0E-20	87.4
SRR943330	12072464	BB-3_1324_1	61.0	59	40	98	6.0E-16	76.3
SRR943330	489755	BA2_13_151 (CX-10)	74.0	50	545	594	3.0E-16	77.4
SRR943330	56563340	BA2_13_151 (CX-10)	74.0	50	545	594	3.0E-16	77.4
SRR943330	45555012	BA2_13_151 (CX-10)	75.0	48	545	592	2.0E-15	74.3
SRR943330	52229171	BA2_13_151 (CX-10)	75.0	48	545	592	2.0E-15	74.3
SRR943331	2447759	JGI24025J20009_100133354	94.0	50	269	318	4.0E-23	100.1
SRR943332	2594970	JGI24025J20009_100133354	93.9	49	264	312	4.0E-22	96.7
SRR943333	15500308	JGI24025J20009_100133354	91.8	49	205	253	2.0E-22	97.8

SRR943333	563618	JGI24025J20009_100133354	70.7	41	8	48	5.0E-12	63.2
SRR943334	48924240	BA2_13_151 (CX-10)	62.3	53	434	486	2.0E-18	84.3
SRR943334	57647517	BA2_13_151 (CX-10)	74.0	50	545	594	3.0E-16	77.4
Pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea, sampled at 5-20 cm bsf (454 GS FLX Titanium)								
SRR363459	8459	CX-10_14344_1	53.1	160	200	357	1.0E-45	175.6
SRR363459	136152	JGI24025J20009_100133354	62.8	51	271	321	4.0E-14	71.6
SRR363460	160142	CD-8_186481_1	54.2	118	12	129	8.0E-35	139.8
SRR363460	232155	BA2_13_151 (CX-10)	50.0	50	19	68	8.0E-10	53.5
SRR363464	139233	CX-10_14344_1	72.7	33	211	243	1.0E-09	56.6
SRR363464	400059	BA1_24_9 (CX-10)	66.1	56	481	536	9.0E-19	86.7
SRR363465	515949	BS-36_20936_1	50.9	59	194	252	2.0E-11	63.2
SRR363497	248437	CX-10_14344_1	40.2	92	36	127	5.0E-16	77.4
Sediment from St. Mandrier harbor near Toulon, France (454 GS FLX Titanium)								
ERR149037	268937	CX-10_14344_1	61.26	111	141	250	2.0E-35	141.4
Hydrothermal vent systems at Axial seamount, northeast Pacific Ocean (Illumina HiSeq 1000)								
ERR420364	2401321	JGI24724J26744_100202833	77.1	35	337	371	5.0E-13	62.8

Table S7.

Metagenomic samples with *Bathyarchaeota* populations (>1%) and/or non-euryarchaeotal *mcrA* genes.

SRA run id	Description	Platform	Novel *mcrA* reads[a]	No. 16S reads	Percentage Bathyarchaeota	Study
SRR001326	Peru Margin subseafloor biosphere, 50 mbsf	454 GS 20	-	16	93.75	(*21*)
SRR001322	Peru Margin subseafloor biosphere, 16 mbsf	454 GS 20	-	51	56.86	(*21*)
SRR001323	Peru Margin subseafloor biosphere, 32 mbsf	454 GS 20	-	61	50.82	(*21*)
SRR001324	Peru Margin subseafloor biosphere, 1 mbsf	454 GS 20	-	33	12.12	(*21*)
SRR363500	Sediments from seabed in the outer part of Oslofjord, Norway, 5-20 cm bsf (sample OF2)	454 GS FLX	-	140	7.86	(*79*)
SRR363460	Sediments from Troll petroleum reservoir in the North Sea, 5-20 cm bsf (sample Tpm1-1)	454 GS FLX	2	96	7.29	(*79*)
ERR476714	Sediments from freshwater lake in Sweden, 10cm bsf	HiSeq 2000	9	5332	7.22	(*80*)
SRR064770	River sediment upstream from a Swedish waste water treatment plant	454 GS FLX	-	34	5.88	(*81*)
SRR363466	Sediments from Troll petroleum reservoir in the North Sea, 5-20 cm bsf (sample Tpm3)	454 GS FLX	-	131	5.34	(*79*)
SRR363499	Sediments from seabed in the outer part of Oslofjord, Norway, 5-20 cm bsf (sample OF2)	454 GS FLX	-	81	4.94	(*79*)
SRR363464	Sediments from Troll petroleum reservoir in the North Sea, 5-20 cm bsf (sample Tpm2)	454 GS FLX	2	120	4.17	(*79*)
ERR149036	Sediment from l'Estaque harbor near Marseille, France, sampled at 2.5m	454 GS FLX	-	153	3.92	-
SRR363498	Sediments from seabed in the outer part of Oslofjord, Norway, 5-20 cm bsf (sample OF1)	454 GS FLX	-	165	3.64	(*79*)
ERR149037	Sediment from St. Mandrier harbor near Toulon, France, sampled at 2.5m bsf	454 GS FLX	1	144	3.47	-
SRR363458	Sediments from Troll petroleum reservoir in the North Sea, 5-20 cm bsf (sample Tplain)	454 GS FLX	-	118	2.54	(*79*)
SRR363497	Sediments from seabed in the outer part of Oslofjord, Norway, 5-20 cm bsf (sample OF1)	454 GS FLX	1	79	2.53	(*79*)
SRR1022349	Sediment from a marine fish farm	HiSeq 2000	7	1261	2.46	(*82*)
SRR948163	Metagenome analysis of temperature-adapted biogas fermenter	Ion Torrent PGM	-	615	1.95	-
SRR363465	Sediments from Troll petroleum reservoir in the North Sea, 5-20 cm bsf (sample Tpm3)	454 GS FLX	1	104	1.92	(*79*)
SRR1291149	Metagenome analysis of temperature-adapted biogas fermenter	Ion Torrent PGM	-	1361	1.91	-
SRR1022378	Estuary sediment	HiSeq 2000	-	1628	1.90	(*82*)
SRR516948	Forest soil metagenome from Breuil-Chenue, France	454 GS FLX	-	172	1.74	-
SRR825189	Arctic soil from Barrow, Alaska, 0.2-0.3 cm	454 GS FLX	-	60	1.67	-
SRR948162	Metagenome analysis of temperature-adapted biogas fermenter	Ion Torrent PGM	-	5173	1.66	-
SRR825190	Arctic soil from Barrow, Alaska, 0.3-0.4 cm	454 GS FLX	-	63	1.59	-
SRR1291147	Metagenome analysis of temperature-adapted biogas fermenter	Ion Torrent PGM	-	13459	1.58	-

SRR389126	Terrestrial mud volcanoes, southwestern Taiwan, 31 cm	454 GS FLX	-	108	0.93	(*83*)
SRR1313285	Tailing pond in Alberta, Canada, 5 mbs	HiSeq 2000	5	7782	0.06	(*84*)
SRR943329	Tailing pond in Alberta, Canada, 6 mbs	HiSeq 2000	2	10546	0.04	(*84*)
SRR1313287	Tailing pond in Alberta, Canada, 5 mbs	HiSeq 2000	3	7233	0.03	(*84*)
SRR1313283	Tailing pond in Alberta, Canada, 5 mbs	HiSeq 2000	3	27791	0.03	(*84*)
SRR943332	Tailing pond in Alberta, Canada, 6 mbs	HiSeq 2000	1	10639	0.02	(*84*)
SRR942941	Tailing pond in Alberta, Canada, 6 mbs	HiSeq 2000	1	10503	0.02	(*84*)
SRR1313281	Tailing pond in Alberta, Canada, 5 mbs	HiSeq 2000	2	5162	0.02	(*84*)
SRR943330	Tailing pond in Alberta, Canada, 6 mbs	HiSeq 2000	5	10639	0.01	(*84*)
SRR943331	Tailing pond in Alberta, Canada , 6 mbs	HiSeq 2000	1	10751	0.01	(*84*)
SRR943334	Tailing pond in Alberta, Canada, 6 mbs	HiSeq 2000	2	10429	0.01	(*84*)
ERR420364	Hydrothermal vent systems at Axial seamount, located in the northeast Pacific Ocean	HiSeq 1000	1	1352	0.00	-
SRR943333	Tailing pond in Alberta, Canada, 6 mbs	HiSeq 2000	2	10639	0.00	(*84*)
SRR099552	Sediment sampled at 0-4 cm bsf from the Tonya Seep, Santa Barbara Channel	454 GS FLX	1	82	0.00	(*85*)
SRR099554	Sediment sampled at 10-15 cm bsf from the Tonya Seep, Santa Barbara Channel	454 GS FLX	1	74	0.00	(*85*)
SRR1313282	Tailing pond in Alberta, Canada, 5 mbs	HiSeq 2000	2	5029	0.00	(*84*)
SRR363459	Sediments from Troll petroleum reservoir in the North Sea, 5-20 cm bsf (sample Tpm1-1)	454 GS FLX	2	62	0.00	(*79*)

[a] Details of sequence similarity to novel Surat Basin or Coal Oil Point *mcrA* genes is given in **Table S6**

Table S8.

Conservation of functionally important Mcr residues identified by Ermler *et al*. (*22*). The functional Mcr particle is composed of two subunits from each α, β and γ subunit, and each active site incorporates residues from both α subunits, which leads to the α' designation for some residues. Residues conserved between known euryarchaeotal methanogens and the BA1 and BA2 genomes are shown in bold.

Residue	Proposed Function	Methanogens	BA1	BA2	Other Surat	Coal Oil Point
α270	CoB interacting[1]	Arg	Gln	Gln	Gln	Asn/Ser/Gly
α481	CoB interacting[2]	Asn	Thr	Ser	Ser/Thr	Thr
α333	CoM interacting[3]	Tyr	Thr	Thr	Thr	Thr/Ala
α443	CoM interacting[4]	Phe	Trp	Trp	Trp	Trp
α271	Methylated[5]	Arg	Arg	Arg	Arg	Arg
α400	Methylated[6]	Gln	His	His	His	His/Ile
α452	Methylated[7]	Cys	Ile	Ile	Ile/Val	Ile/Val
α330	Substrate cavity wall	Phe	Phe	Phe	Phe	Phe
α445	Thiol substituted	Gly	Gly	Gly	Gly	Gly
α'225	CoB interacting	Arg	Arg	Arg	Arg	Arg
α'256	CoB interacting	Lys	Lys	Lys	Lys	Lys/Arg
α'257	CoB interacting, methylated	His	His	His	His	His
α'147	F_{430} axial ligand	Gln	Gln	Gln	Gln	Gln
β379	CoB interacting	His	His	His		
β367	CoM interacting	Tyr	Tyr	Tyr		
β361	Substrate cavity wall	Phe	Phe	Phe		
γ120	CoM interacting	Arg	Arg	Arg		

[1]In the absence of arginine, the amine group of glutamine could still donate a hydrogen bond to one of the phosphate oxygen of CoB for the Surat *Bathyarchaeota* McrA sequences (**Fig. S11**).
[2]Asparagine, threonine, and serine all have polar, uncharged side-chains.
[3]Tyrosine and threonine both have hydroxyl groups and polar, uncharged side-chains.
[4]Phenylalanine and tryptophan are aromatic amino acids with hydrophobic, nonpolar, uncharged side-chains.
[5]Identified as being methylated across all methanogens (*86*).
[6]Not methylated in the crystal structure of *Methanosarcina bakeri* (*86*, *87*).
[7]Not always methylated (*86*).

Table S9.

Annotated genes used for metabolic reconstruction of the BA1 genome.

Contig ID	Symbol	Gene Product
Methanogenesis		
15_62 45_25 89_45	*mtrH*	Methyl-H_4MPT reductase subunit H
15_63 45_24 89_47	*mttB, mtbB*	Corrinoid proteins associated with *mtrH*
45_16	*mttC*	Trimethylamine methyltransferase
5_57 45_13 87_11	*mtbC*	Dimethylamine methyltransferase
45_14	*mtbA*	Methylcobamide:CoM methyltransferase
45_23	*mtsA*	Methylated-thiol-coenzyme M methyltransferase
45_18	*mtaA*	Methylcobalamin: coenzyme M methyltransferase
24_09	*mcrA*	Methyl-coenzyme M reductase subunit A
24_07	*mcrB*	Methyl-coenzyme M reductase subunit B
24_08	*mcrG*	Methyl-coenzyme M reductase subunit G
3_22 3_36 87_38	*hdrD*	Heterodisulfide reductase subunit D
3_23 3_37 87_39	*glcD*	Dehydrogenase related to lactate dehydrogenases
66_57	*mvhA*	Methylviologen reducing hydrogenase subunit A
66_58	*mvhG*	Methylviologen reducing hydrogenase subunit G
70_36 66_61	*mvhD*	Methylviologen reducing hydrogenase subunit D
85_4 87_34	*mvhB*	Methylviologen reducing hydrogenase subunit B
42_5 52_9 85_28 66_63	*hdrA*	Heterodisulfide reductase subunit A
10_11 52_6 66_59 85_26	*hdrB*	Heterodisulfide reductase subunit B
10_12 52_5 66_60 85_27 87_37	*hdrC*	Heterodisulfide reductase subunit C
15_40 60_1 7_30 81_10 81_9 6_16 88_26 88_27 89_15 89_21 89_23 26_9	*met*	Undefined methyltransferases
51_19	*fae/hps*	Bifunctional enzyme Fae/Hps
2_1	*echA*	Energy converting hydrogenase subunit A
2_8	*echB*	Energy converting hydrogenase subunit B
2_17	*echC*	Energy converting hydrogenase subunit C
2_18	*echD*	Energy converting hydrogenase subunit D
2_19	*echE*	Energy converting hydrogenase subunit E
2_20	*echG*	Energy converting hydrogenase subunit G
Reductive acetyl-CoA pathway		
70_31	*CooS*	Carbon monoxide dehydrogenase
70_29	*CooC*	Carbon monoxide dehydrogenase accessory protein
89_54	*fwdA*	Tungsten-containing formylmethnaofuran dehydrogenase subunit A
89_55	*fwdB*	Tungsten-containing formylmethnaofuran dehydrogenase subunit B
89_53	*fwdC*	Tungsten-containing formylmethnaofuran dehydrogenase subunit C
89_56	*fwdD*	Tungsten-containing formylmethnaofuran dehydrogenase subunit D
58_8 89_62	*ftr*	Formylmethanofuran-H4MPT formyltransferase
89_48	*mch*	Methenyl-H_4MPT cyclohydrolase
89_61	*mtd*	Methylene-H_4MPT dehydrogenase
25_3 44_18 81_11	*mer*	5,10-methylene-H_4MPT reductase
70_37	*frhA*	F_{420}-reducing hydrogenase subunit A
89_31	*frhB*	F_{420}-reducing hydrogenase subunit B
70_38	*frhG*	F_{420}-reducing hydrogenase subunit G
5_15	*cdhA*	Acetyl-CoA decarbonylase/synthase complex subunit A
5_3	*cdhB*	Acetyl-CoA decarbonylase/synthase complex subunit B
5_2	*cdhD*	Acetyl-CoA decarbonylase/synthase complex subunit D
5_4	*cdhE*	Acetyl-CoA decarbonylase/synthase complex subunit E

5_1	*cdhG*	Acetyl-CoA decarbonylase/synthase complex subunit G
2_34	*porA*	Pyruvate: Ferredoxin oxidoreductase subunit A
2_35	*porB*	Pyruvate: Ferredoxin oxidoreductase subunit B
2_33	*porD*	Pyruvate: Ferredoxin oxidoreductase subunit D
2_32	*porG*	Pyruvate: Ferredoxin oxidoreductase subunit G
33_7 57_3 87_14	*acd*	ADP-forming acetyl-CoA synthase
Glycolysis/Gluconeogenesis		
45_20	GH38	Alpha-mannosidase
20_4	*fbp*	Fructose-1,6-bisphosphatase
51_38	*tpiA*	Triosephosphate isomerase
7_54	*gspA*	Glycerol-3-phosphate dehydrogenase
66_22	*pgiA*	Glucose-6-phosphate isomerase
69_8	*gap*	Glyceraldehyde-3-phosphate dehydrogenase
66_7	*pgk*	Phosphoglycerate kinase
51_5	*pgm*	2,3-bisphosphoglycerate-independent phosphoglycerate mutase
66_15	*gyd*	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
51_10 53_6 17_23 89_30 87_36	*pks*	Pyruvate kinase, alpha/beta domain
53_8	*maeA*	NAD-dependent malic enzyme
6_20	*pmi*	Bifunctional phosphoglucose/phosphomannose isomerase
Urea Cycle		
76_22	*cpkA*	Carbamate kinase
82_1	*purA*	Adenylosuccinate synthetase
3_12 41_11 68_9 84_39	*argF*	Ornithine carbamoyltransferase
16_19	*pyrB*	Aspartate carbamoyltransferase
63_8	*pdaD*	Pyruvoyl-dependent arginine decarboxylase
16_61	*speB*	Agmatinase
Tricarboxylic acid cycle		
58_6	*mdh*	Malate dehydrogenase
9_6	*fum*	Fumarate hydratase
9_7	*sdhA*	Succinate dehydrogenase subunit A
9_8	*sdhB*	Succinate dehydrogenase subunit B
15_23 67_9	*scsA*	Succinyl-CoA synthase subunit A
15_22	*scsB*	Succinyl-CoA synthase subunit B
42_3	*korA*	2-oxoglutarate: ferredoxin oxidoreductase subunit A
42_2	*korB*	2-oxoglutarate: ferredoxin oxidoreductase subunit B
42_1	*korC*	2-oxoglutarate: ferredoxin oxidoreductase subunit C
71_16	*pycA*	Pyruvate carboxylase subunit A
71_17	*pycB*	Pyruvate carboxylase subunit B
Calvin Cycle/pentose phosphate pathway		
13_71 13_73		Ribose 1,5-bisphosphate isomerase
13_72	*prs*	Ribose-phosphate pyrophosphokinase
13_70	*rbcL*	Ribulose bisphosphate carboxylase
15_32	*rpiA*	Ribose-5-phosphate isomerase A
FA synthesis/b-oxidation		
4_32	*fadD*	Long chain fatty acid-CoA ligase
5_61	*fixB*	Electron transferring flavoprotein subunit B
5_62	*fixA*	Electron transferring flavoprotein subunit A
87_28	*fadA*	3-ketoacyl-CoA thiolase
Amino acid fermentation		
22_3 53_19 58_7 62_11 16_122	*aspC*	Aspartate aminotransaferase
13_69	*bkdA*	Branched chain amino acid dehydrogenase complex E1
13_68	*bkdB*	Branched chain amino acid dehydrogenase complex E2

13_67	*bkdC*	Branched chain amino acid dehydrogenase complex E3
46_14	*vorA*	Isovalerate:ferredoxin oxidoreductase subunit A
46_15	*vorB*	Isovalerate:ferredoxin oxidoreductase subunit B
46_16	*vorC*	Isovalerate:ferredoxin oxidoreductase subunit C
46_13	*vorG*	Isovalerate:ferredoxin oxidoreductase subunit G
42_3	*korA*	Ketoglutarate:ferredoxin oxidoreductase subunit A
42_2	*korB*	Ketoglutarate:ferredoxin oxidoreductase subunit B
42_1	*korC*	Ketoglutarate:ferredoxin oxidoreductase subunit C
71_78	*mmdA*	Methmalonyl-CoA carboxyltransferase
87_7	*epi*	Methmalonyl-CoA epimerase
5_58 87_9	*mutC*	Methmalonyl-CoA mutase
76_16	*mmsD*	Methylmalonate semialdehyde dehydrogenase
50_3	*iorA*	Indolepyruvate:ferredoxin oxidoreductase subunit A
50_4	*iorB*	Indolepyruvate:ferredoxin oxidoreductase subunit B
43_12	*gcvP*	Glycine dehydrogenase (decarboxylating) P subunit
69_10	*gcvH*	Glycine cleavage system H
36_7 36_8	*gcvP1*	Glycine cleavage system P1
36_4 36_5 36_6	*gcvP2*	Glycine cleavage system P2
3_31	*gdh*	Glutamate dehydrogenase
47_16	*grdA*	Glycine/sarcosine/betaine reductase complex subunit A
14_15	*grdB*	Glycine/sarcosine/betaine reductase complex subunit B
47_14	*grdC*	Glycine/sarcosine/betaine reductase complex subunit C
6_42	*hutH*	Histidine ammonia lyase
6_38	*hutU*	Urocanate hydratase
6_37 26_8	*hutI*	Imidazolonenpropionase
28_10 3_9 65_42 81_15 66_34 66_48 87_33	*aor*	Tungsten-containing aldehyde ferredoxin oxidoreductase
18_3	*for*	Tungsten-containing formaldehyde: ferredoxin oxidoreductase
Amino acid synthesis		
9_49 3_11 13_22 42_16 17_55 68_11 84_54 16_168	*thrC*	Threonine synthase
15_53 63_11 16_147	*trpB*	Tryptophan synthase beta chain
28_7	*trpA*	Tryptophan synthase alpha chain
16_133	*ansAB*	L-asparaginase
21_7 13_10 13_11 72_13 66_12 66_27	*tdh*	L-theronine dehydrogenase
49_3 16_138	*glyA*	Serine hydroxymethyltransferase
13_74	*pgdh*	3-Phosphoglycerate dehydrogenase
50_5	*leuD*	Isopropylmalate/citramalate isomerase small subunit
50_6	*leuC*	Isopropylmalate/citramalate isomerase large subunit
71_20	*leuB*	3-isopropylmalate/3-methylmalate dehydrogenase
6_58	*cysA*	Cysteine synthase A
85_15	*glnA*	Glutamine synthetase
68_15	*metE*	Methionine synthase
31_4	*lysA*	Diaminopimelate decarboxylase
31_5	*dapF*	Diaminopimelate epimerase
33_4	*asnB*	Asparagine synthetase B
16_1	*serB*	Phosphoserine phosphatase
Transporters		
71_2	*oppA*	Branched chain amino acid
71_3	*oppB*	Branched chain amino acid
71_4	*oppC*	Branched chain amino acid
71_5	*oppD*	Branched chain amino acid

71_6	*oppE*	Branched chain amino acid
12_28	*appF*	Oligopeptides transporter
12_29	*appE*	Oligopeptides transporter
12_30	*appD*	Oligopeptides transporter
12_31	*appB*	Oligopeptides transporter
12_32	*appA*	Oligopeptides transporter
32_14 41_17 50_11	*dppA*	Dipeptide transporter
32_15 41_18 50_12	*dppB*	Dipeptide transporter
32_16 41_19 50_13	*dppD*	Dipeptide transporter
41_20		Ni^{2+} transporter
41_21		Ni^{2+} transporter
82_2		Fe^{3+} transporter
82_3		Fe^{3+} transporter
82_4		Fe^{3+} transporter
65_14		Co^{2+} transporter
65_13		Co^{2+} transporter
65_12		Co^{2+} transporter
70_4	*safE*	Sulfite exporter
62_5	*potA*	Spermidine transporter
62_6		Solute transporter associated with *potA*
62_7		Solute transporter associated with *potA*
66_28		Phosphate transporter
45_3	malF	Maltose transporter
45_4	malG	Maltose transporter
45_5	malK	Maltose transporter
5_31	*natA*	Na^+ transporter
5_32	*natB*	Na^+ transporter
2_21	*nha*	Na^+/H^+ antiporter
16_144		K^+ transporter
16_145		K^+ transporter
16_146		K^+ transporter
11_16	*hppA*	K^+/H^+ antiporter
11_15	*mgtA*	K^+/H^+ antiporter
6_55		Na^+/solute symporter
Peptidases		
41_22 52_17 57_2 80_13	M50	Peptidase M50
7_34 15_37 50_10	M29	Thermophillic metaloprotease
32_9	M48	Peptidase M48
66_35 14_1	M20	Peptidase M20
84_55	S58	Peptidase S58
6_60 66_33	*pepQ*	Xaa-Pro dipeptidase
38_23	S9	Peptidase S9
80_31	*map*	Methionine aminopeptidase
Cofactor biosynthesis		
13_25	*comB*	2-phosphosulfolactate phosphatase
89_14	*mptN*	Tetrahydromethanopterin:alph-L-glutamate ligase
11_1 11_28 44_8 13_31 17_39 48_31 2_40	*moaA*	Cyclic pyranopterin monophosphate synthase
89_12	*mobA*	Molybdenum cofactor guanylyltransferase
17_40	*moaB*	Molybdopterin adenylyltransferase
17_41	*moaC*	Cyclic pyranopterin monophosphate synthase accessory protein
42_18	*moaE*	Molybdpterin guanine dinucleotide biosynthesis protein
89_2	*cofE*	Coenzyme F_{420}-0:L-glutamate ligase
13_3 13_4	*moeA*	Molybdpterin biosynthesis protein

85_20	*hypA*	Hydrogenase nickel incorporation protein A
85_21 5_56	*hypB*	Hydrogenase nickel incorporation protein B
85_22	*hypC*	Hydrogenase nickel incorporation protein C
85_16	*hypD*	Hydrogenase nickel incorporation protein D
85_5	*hypE*	Hydrogenase nickel incorporation protein E
85_25	*hypF*	Hydrogenase nickel incorporation protein F
16_167	*thiDN*	Bifunctional thiamine biosynthesis protein
16_166	*thiW*	Thiamine-precursor transporter protein
38_3	*thiC*	Thiamine biosynthesis protein
46_22	*hemD*	Uroporphyrinogen-III synthase
46_24	*cobA*	Uroporphyrinogen-III C-methyltransferase
69_14	*bioB*	Biotin synthase
38_8	*iscS*	Cysteine desulfurase

Table S10.

Annotated genes used for metabolic reconstruction of the BA2 genome.

Contig ID	Gene	Gene Product
Methanogenesis		
13_151	*mcrA*	Methyl-coenzyme M reductase subunit A
13_159	*mcrB*	Methyl-coenzyme M reductase subunit B
13_160	*mcrG*	Methyl-coenzyme M reductase subunit G
13_138	*mcrC*	Methyl-coenzyme M reductase subunit C
13_139	*mcrD*	Methyl-coenzyme M reductase subunit D
13_158	*atwA*	Methyl-coenzyme M reductase subunit A2
43_1 13_135 13_147	*hdrD*	Heterodisulfide reductase subunit D
13_146 13_136	*glcD*	Dehydrogenase related to lactate dehydrogenases
22_26	*mvhA*	Methylviologen reducing hydrogenase subunit A
22_27	*mvhG*	Methylviologen reducing hydrogenase subunit G
22_32	*mvhD*	Methylviologen reducing hydrogenase subunit D
22_20	*mvhB*	Methylviologen reducing hydrogenase subunit B
22_47	*hdrA/mvhD*	*hdrA*/*mvhD* fusion protein
15_11 22_31 22_34	*hdrA*	Heterodisulfide reductase subunit A
22_28 13_121	*hdrB*	Heterodisulfide reductase subunit B
22_29 13_122	*hdrC*	Heterodisulfide reductase subunit C
10_31 27_11 4_31	*met*	Undefined methyltransferases
21_10 40_12	*fae/hps*	Bifunctional enzyme *fae*/*hps*
Reductive acetyl-CoA pathway		
10_56	*ftr*	Formylmethanofuran-H4MPT formyltransferase
55_15	*cdhA*	Acetyl-CoA decarbonylase/synthase complex subunit A
55_17	*cdhB*	Acetyl-CoA decarbonylase/synthase complex subunit B
55_18	*cdhD*	Acetyl-CoA decarbonylase/synthase complex subunit D
55_16	*cdhE*	Acetyl-CoA decarbonylase/synthase complex subunit E
55_19	*cdhG*	Acetyl-CoA decarbonylase/synthase complex subunit G
51_23	*porA*	Pyruvate: Ferredoxin oxidoreductase subunit A
51_24	*porB*	Pyruvate: Ferredoxin oxidoreductase subunit B
51_22	*porD*	Pyruvate: Ferredoxin oxidoreductase subunit D
51_21	*porG*	Pyruvate: Ferredoxin oxidoreductase subunit G
17_82 34_21 19_35	*acd*	ADP-forming acetyl-CoA synthase
4_27	*acs*	Acetyl-CoA synthase
Glycolysis/Gluconeogenesis		
19_48	*pfk*	Fructose-1,6-bisphosphatase
19_36	*tpiA*	Triosephosphate isomerase
17_34	*pgiA*	Glucose-6-phosphate isomerase
50_11	*pgk*	Phosphoglycerate kinase
3_8 19_59	*pks*	Pyruvate kinase, alpha/beta domain
11_36	*pck*	Phosphoenolpyruvate carboxykinase
Urea Cycle		
6_45	*purA*	Adenylosuccinate synthetase
5_12	*asl*	Aginosuccinate lyase
5_49 26_30	*argF*	Ornithine carbamoyltransferase
1_3	*pyrB*	Aspartate carbamoyltransferase
4_3	*pdaD*	Pyruvoyl-dependent arginine decarboxylase
26_16	*speE*	Spermidine synthase
5_10	*argG*	Arginosuccinate synthase
16_22	*argD*	Acetyl ornithine aminotransferase

16_21	*argB*	Acetylglutamate kinase
16_20 5_11	*argC*	N-acetyl-gamma-glutamyl-phosphate reductase
Tricarboxylic acid cycle		
34_8 34_9	*fum*	Fumarate hydratase
Calvin Cycle/pentose phosphate pathway		
6_30	*phi*	3-hexulose-6-phosphate isomerase
FA synthesis/b-oxidation		
31_16	*fadB*	3-hydroxyacyl-CoA dehydrogenase
43_5	*fixB*	Electron transferring flavoprotein subunit B
43_4	*fixA*	Electron transferring flavoprotein subunit A
2_33	*fadA*	3-ketoacyl-CoA thiolase
21_3 44_1 44_2	*caiA*	Acyl-CoA dehydrogenase
12_53 17_46	*ech*	Enoyl-CoA hydratase
21_1	*caiC*	Fatty acid-CoA ligase
Amino acid fermentation		
9_6	*phnW*	2-aminoehtylphophonate-pyruvate transaminase
50_13	*gcvH*	Glycine cleavage system H
22_19 13_129 13_161 13_162 13_163 13_164	*aor*	Tungsten-containing aldehyde ferredoxin oxidoreductase
13_41	*for*	Tungsten-containing formaldehyde:ferredoxin oxidoreductase
Amino acid synthesis		
9_7	*serA*	Phosphoglycerate dehydrogenase
6_31 24_39	*lysC*	Aspartate kinase
6_32	*thrA*	Homoserine dehydrogenase
6_34	*thrB*	Homoserine kinase
6_33 12_49 54_4	*thrC*	Threonine synthase
13_83	*ilvD*	Dihydroxy-acid dehydrogenase
13_84	*cimA*	(R)-citamalate synthase
13_85	*ilvB*	Acetolactate synthase large subunit
13_86	*ilvH*	Acetolactate synthase small subunit
13_87	*ilvC*	Ketol-acid reductoisomerase
24_40	*asd*	Aspartate-semialdehyde dehydrogenase
24_41	*aroA'*	2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase
24_42	*aroB'*	3-dehydroquinate synthase
24_43	*aroD*	3-dehydroquinate dehydratase
24_44	*aroE*	Shikimate dehydrogenase
24_45	*aroK*	Shikimate kinase
24_46	*aroA*	3-phosphoshikimate 1-carboxyvinyltransferase
24_47	*aroC*	Chorismate synthase
24_48	*pheA1*	Chorismate mutase
24_50	*tyrA*	Prephenate dehydrogenase
24_55 24_60	*pheA2*	Prephenate dehydratase
24_49 17_24	*aspC*	Aspartate aminotransferase
17_19	*hisI*	Imidazole glycerol phosphate synthase
17_20 17_21	*hisF*	Imidazole glycerol phosphate synthase
17_22	*hisB*	Imidazole glycerol phosphate dehydratase
17_25	*hisD*	Histidinol dehydrogenase
17_26	*hisG*	ATP phosphoribosyltransferase
20_10	*trpE*	Anthranilate synthase component 1
20_09	*trpG*	Anthranilate synthase component 2
20_08	*trpD*	Anthranilate phosphoribosyltransferase
20_07	*trpC*	Indole-3-glycerol phosphate synthase
20_06	*trpF*	N-(5'-phosphoribosyl)anthranilate isomerase
17_45 20_05	*trpB*	Tryptophan synthase beta chain

20_04	*trpA*	Tryptophan synthase alpha chain
5_32	*leuA*	2-isopropylmalate synthase
23_28 5_31	*leuB*	3-isopropylmalate/3-methylmalate dehydrogenase
5_29	*leuC*	Isopropylmalate/citramalate isomerase large subunit
5_30	*leuD*	Isopropylmalate/citramalate isomerase small subunit
40_2	*glyA*	Serine-hydroxymethyltransferase
13_1	*glnA*	Glutamine synthetase
11_30	*metE*	Methionine synthase
8_22 20_22	*asnB*	Asparagine synthetase B
1_26	*serB*	Phosphoserine phosphatase
10_69	*proC*	Pyrroline-5-carboxylate reductase
Transporters		
22_8		Cobalt transporter
22_9		Cobalt transporter
22_10		Cobalt transporter
13_127	*btuD*	Cobalamin import ATP-binding protein
13_126	*btuC*	Cobalamin import system permease protein
13_125	*btuF*	Cobalamin-binding protein precursor
34_24	*tauA*	Bicarbonate transporter
34_25	*tauB*	Bicarbonate transporter
9_9	*natA*	Na^+ transporter
9_12	*natB*	Na^+ transporter
50_10	*nha*	Na^+/H^+ antiporter
1_24 1_25		K^+ transporter
31_3	*copA*	K^+/H^+ antiporter
31_2	*hppA*	K^+/H^+ antiporter
Peptidases		
4_30 10_30	M50	Peptidase M50
19_29	M48	peptidase M48
32_5	*pepQ*	Xaa-Pro dipeptidase
19_20	*map*	Methionine aminopeptidase
Cofactor biosynthesis		
17_71	*mptN*	Tetrahydromethanopterin:alph-L-glutamate ligase
51_35	*mptB*	Dihydroneopterin 2',3'-cyclic phosphate phosphodiesterase
8_3	*moaA*	Cyclic pyranopterin monophosphate synthase
8_1	*moaB*	Molybdopterin adenylyltransferase
8_2	*moaC*	Cyclic pyranopterin monophosphate synthase accessory protein
51_33	*moaE*	Molybdopterin guanine dinucleotide biosynthesis protein
35_13	*cofE*	Coenzyme F_{420}-0:L-glutamate ligase
22_51	*hypA*	Hydrogenase nickel incorporation protein A
22_50	*hypB*	Hydrogenase nickel incorporation protein B
22_49	*hypC*	Hydrogenase nickel incorporation protein C
22_55	*hypD*	Hydrogenase nickel incorporation protein D
22_48	*hypF*	Hydrogenase nickel incorporation protein F
11_7	*thiDN*	Bifunctional thiamine biosynthesis protein
4_23	*iscS*	Cysteine desulfurase
21_11	*mfnA*	Tyrosine decarboxylase
Other		
24_16 - 24_34	*gvpA-O*	Gas vesicle proteins
22_25	*shyA*	Sulfohydrogenase 2 subunit A
22_24	*shyD*	Sulfohydrogenase 2 subunit D
22_23	*shyG*	Sulfohydrogenase 2 subunit G
22_22	*shyB*	Sulfohydrogenase 2 subunit B

Table S11.

Percentage of reads in CX-10 mapping to assembled scaffolds and the BA1 and BA2 genomes.

Pairs	# pairs	% pairs	% QC pairs	% assembly	% genomes
Total reads in CX-10	44,642,874	100	-	-	-
Quality checked	40,620,903	91.0	100	-	-
Mapped to assembled sequence ≥500 bp	21,124,518	47.3	52.0	100	-
Mapped to population genomes	10,059,952	22.5	24.8	47.6	100
Mapped to BA1 genome	372,154	0.83	0.92	1.76	3.70
Mapped to BA2 genome	418,410	0.94	1.03	1.98	4.16

Table S12.

Characteristics of 16S rRNA genes within the BA1 and BA2 genomes.

	BA1	BA2
Gene length (bp) [a]	1,516	1,677
Scaffold length (bp)	34,298	43,385
Start position in contig	4	33,820
End position in contig	1,519	35,496
Best match in Greengenes[b]	c__MCG	c__MCG; o__B10
Accession	2078987	561428
Alignment length	1499 of 1502	1328 of 1468
Percent identity (%)	91.1%	95.0%
Best match in NR[d]	uncultured *Desulfurococcales*[c]	uncultured archaeon
Accession	HQ700684.1	AB213058.1
Alignment length	1366 of 1499	1469 of 1559
Percent identity (%)	91%	94%
Best match in RefSeq RNA[d]	*Thermofilum pendens* Hvv 3	*Thermofilum pendens* Hrk 5
Accession	NR_029214.1	NR_074406.1
Alignment length	1267 of 1530	1151 of 1365
Percent identity (%)	83%	84%

[a] 16S rRNA genes identified with the SSU finder function of CheckM v1.0.0 (*28*)
[b] BLASTN v2.2.26+ with default parameters against May, 2013 database
[c] Identical to sequence within the Greengenes database
[d] NCBI web service with default parameters on May 1, 2015

Table S13.

Marker genes used to infer archaeal genome tree.

HMM Identifier	Description	Model length
PF00181	Ribosomal Proteins L2, RNA binding domain	77
PF00281	Ribosomal protein L5	56
PF00298	Ribosomal protein L11, RNA binding domain	69
PF00312	Ribosomal protein S15	83
PF00334	Nucleoside diphosphate kinase	135
PF00410	Ribosomal protein S8	129
PF00466	Ribosomal protein L10	100
PF00673	Ribosomal L5P family C-terminus	95
PF00687	Ribosomal protein L1p/L10e family	220
PF00832	Ribosomal L39 protein	43
PF00861	Ribosomal L18p/L5e family	119
PF00900	Ribosomal family S4e	77
PF00935	Ribosomal protein L44	77
PF01000	RNA polymerase Rpb3/RpoA insert domain	112
PF01015	Ribosomal S3Ae family	195
PF01090	Ribosomal protein S19e	140
PF01092	Ribosomal protein S6e	127
PF01157	Ribosomal protein L21e	99
PF01191	RNA polymerase Rpb5, C-terminal domain	74
PF01192	RNA polymerase Rpb6	57
PF01194	RNA polymerases N / 8 kDa subunit	60
PF01198	Ribosomal protein L31e	83
PF01200	Ribosomal protein S28e	69
PF01246	Ribosomal protein L24e	71
PF01269	Fibrillarin	229
PF01280	Ribosomal protein L19e	148
PF01655	Ribosomal protein L32	110
PF01667	Ribosomal protein S27	55
PF01798	Putative snoRNA binding domain	150
PF01864	Putative integral membrane protein DUF46	175
PF01868	Domain of unknown function UPF0086	89
PF01922	SRP19 protein	95
PF01949	Protein of unknown function DUF99	187
PF01951	Archease protein family (MTH1598/TM1083)	137
PF01984	Double-stranded DNA-binding domain	107
PF01994	tRNA ribose 2'-O-methyltransferase, aTrm56	121
PF02006	Protein of unknown function DUF137	178
PF02978	Signal peptide binding domain	104
PF03874	RNA polymerase Rpb4	117
PF03911	Sec61beta family	41
PF03946	Ribosomal protein L11, N-terminal domain	60
PF03947	Ribosomal Proteins L2, C-terminal domain	130
PF04010	Protein of unknown function (DUF357)	75
PF04019	Protein of unknown function (DUF359)	121
PF04032	RNAse P Rpr2/Rpp21/SNM1 subunit domain	85
PF04414	D-aminoacyl-tRNA deacylase	214
PF04561	RNA polymerase Rpb2, domain 2	190
PF04563	RNA polymerase beta subunit	203
PF04565	RNA polymerase Rpb2, domain 3	68
PF04919	Protein of unknown function (DUF655)	181
PF07541	Eukaryotic translation initiation factor 2 alpha subunit	114
PF08068	DKCLD (NUC011) domain	59

PF08069	Ribosomal S13/S15 N-terminal domain	60
PF08071	RS4NT (NUC023) domain	38
PF13507	CobB/CobQ-like glutamine amidotransferase domain	259
PF13656	RNA polymerase Rpb3/Rpb11 dimerisation domain	77
TIGR00012	L29: ribosomal protein L29	56
TIGR00021	rpiA: ribose 5-phosphate isomerase A	218
TIGR00037	eIF_5A: translation elongation factor IF5A	130
TIGR00042	TIGR00042: non-canonical purine NTP pyrophosphatase, RdgB/HAM1 family	184
TIGR00064	ftsY: signal recognition particle-docking protein FtsY	279
TIGR00111	pelota: mRNA surveillance protein pelota	351
TIGR00134	gatE_arch: glutamyl-tRNA(Gln) amidotransferase, subunit E	622
TIGR00240	ATCase_reg: aspartate carbamoyltransferase, regulatory subunit	150
TIGR00264	TIGR00264: alpha-NAC homolog	116
TIGR00270	TIGR00270: TIGR00270 family protein	154
TIGR00279	L10e: ribosomal protein L10.e	172
TIGR00280	L37a: ribosomal protein L37a	92
TIGR00283	arch_pth2: peptidyl-tRNA hydrolase	115
TIGR00291	RNA_SBDS: rRNA metabolism protein, SBDS family	231
TIGR00296	TIGR00296: uncharacterized protein, PH0010 family	200
TIGR00307	S8e: ribosomal protein S8.e	127
TIGR00308	TRM1: N2,N2-dimethylguanosine tRNA methyltransferase	375
TIGR00323	eIF-6: putative translation initiation factor eIF-6	215
TIGR00337	PyrG: CTP synthase	526
TIGR00373	TIGR00373: transcription factor E	162
TIGR00392	ileS: isoleucine--tRNA ligase	861
TIGR00398	metG: methionine--tRNA ligase	530
TIGR00405	L26e_arch: ribosomal protein L24	145
TIGR00408	proS_fam_I: proline--tRNA ligase	475
TIGR00419	tim: triose-phosphate isomerase	228
TIGR00422	valS: valine--tRNA ligase	863
TIGR00425	CBF5: putative rRNA pseudouridine synthase	322
TIGR00432	arcsn_tRNA_tgt: archaeosine tRNA-ribosyltransferase	637
TIGR00442	hisS: histidine--tRNA ligase	406
TIGR00448	rpoE: DNA-directed RNA polymerase	179
TIGR00456	argS: arginine--tRNA ligase	569
TIGR00463	gltX_arch: glutamate--tRNA ligase	560
TIGR00468	pheS: phenylalanine--tRNA ligase, alpha subunit	324
TIGR00471	pheT_arch: phenylalanine--tRNA ligase, beta subunit	551
TIGR00489	aEF-1_beta: translation elongation factor aEF-1 beta	88
TIGR00490	aEF-2: translation elongation factor aEF-2	720
TIGR00491	aIF-2: translation initiation factor aIF-2	594
TIGR00521	coaBC_dfp: phosphopantothenoylcysteine decarboxylase / phosphopantothenate--cysteine ligase	392
TIGR00522	dph5: diphthine synthase	258
TIGR00549	mevalon_kin: mevalonate kinase	276
TIGR00581	moaC: molybdenum cofactor biosynthesis protein C	148
TIGR00729	TIGR00729: ribonuclease HII	207
TIGR00877	purD: phosphoribosylamine--glycine ligase	425
TIGR00878	purM: phosphoribosylformylglycinamidine cyclo-ligase	332
TIGR00884	guaA_Cterm: GMP synthase (glutamine-hydrolyzing), C-terminal domain	310
TIGR00982	S23_S12_E_A: ribosomal protein S23 (S12)	139
TIGR01008	rpsC_E_A: ribosomal protein S3	195
TIGR01012	Sa_S2_E_A: ribosomal protein S2	196
TIGR01018	rpsD_arch: ribosomal protein S4	162
TIGR01020	rpsE_arch: ribosomal protein S5	212
TIGR01025	rpsS_arch: ribosomal protein S19	135
TIGR01028	S7_S5_E_A: ribosomal protein S7	186
TIGR01038	L22_arch: ribosomal protein L22	148
TIGR01046	S10_Arc_S20_Euk: ribosomal protein S10	99
TIGR01077	L13_A_E: ribosomal protein L13	141

TIGR01080	rplX_A_E: ribosomal protein L24	116
TIGR01177	TIGR01177: TIGR01177 family protein	330
TIGR01213	pseudo_Pus10arc: tRNA pseudouridine(54/55) synthase	387
TIGR01309	L30P_arch: ribosomal protein L30P	151
TIGR01578	MiaB-like-B: MiaB-like tRNA modifying enzyme, archaeal-type	421
TIGR01736	FGAM_synth_II: phosphoribosylformylglycinamidine synthase II	718
TIGR02076	pyrH_arch: putative uridylate kinase	222
TIGR02153	gatD_arch: glutamyl-tRNA(Gln) amidotransferase, subunit D	405
TIGR02338	gimC_beta: prefoldin, beta subunit	110
TIGR02389	RNA_pol_rpoA2: DNA-directed RNA polymerase, subunit A"	367
TIGR02390	RNA_pol_rpoA1: DNA-directed RNA polymerase subunit A'	867
TIGR03626	L3_arch: archaeal ribosomal protein L3	331
TIGR03627	arch_S9P: archaeal ribosomal protein S9P	130
TIGR03628	arch_S11P: archaeal ribosomal protein S11P	117
TIGR03629	arch_S13P: archaeal ribosomal protein S13P	144
TIGR03630	arch_S17P: archaeal ribosomal protein S17P	102
TIGR03636	L23_arch: archaeal ribosomal protein L23	77
TIGR03653	arch_L6P: archaeal ribosomal protein L6P	170
TIGR03665	arCOG04150: arCOG04150 universal archaeal KH domain protein	173
TIGR03670	rpoB_arch: DNA-directed RNA polymerase subunit B	599
TIGR03671	cca_archaeal: CCA-adding enzyme	410
TIGR03672	rpl4p_arch: 50S ribosomal protein L4P	251
TIGR03673	rpl14p_arch: 50S ribosomal protein L14P	131
TIGR03677	rpl7ae: 50S ribosomal protein L7Ae	117
TIGR03679	arCOG00187: universal archaeal metal-binding-domain/4Fe-4S-binding-domain containing ABC transporter, ATP-binding protein	218
TIGR03680	eif2g_arch: translation initiation factor 2, gamma subunit	407
TIGR03682	arCOG04112: diphthamide biosynthesis enzyme Dph2	308
TIGR03683	A-tRNA_syn_arch: alanine--tRNA ligase	902
TIGR03684	arCOG00985: arCOG04150 universal archaeal PUA-domain protein	154
TIGR03685	L12P_arch: 50S ribosomal protein L12P	105
TIGR03722	arch_KAE1: universal archaeal protein Kae1	323
TIGR03724	arch_bud32: Kae1-associated kinase Bud32	199
TIGR03800	PLP_synth_Pdx2: pyridoxal 5'-phosphate synthase, glutaminase subunit Pdx2	184

REFERENCES AND NOTES

1. W. Reeburgh, Oceanic methane biogeochemistry. *Chem. Rev.* **107**, 486–513 (2007). Medline doi:10.1021/cr050362v

2. R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, R. Hedderich, Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **6**, 579–591 (2008). Medline doi:10.1038/nrmicro1931

3. R. K. Thauer, Anaerobic oxidation of methane with sulfate: On the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO_2. *Curr. Opin. Microbiol.* **14**, 292–299 (2011). Medline doi:10.1016/j.mib.2011.03.003

4. S. Gribaldo, C. Brochier-Armanet, The origin and evolution of Archaea: A state of the art. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1007–1022 (2006). Medline doi:10.1098/rstb.2006.1841

5. S. Kelly, B. Wickstead, K. Gull, Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc. Biol. Sci.* **278**, 1009–1018 (2011). Medline doi:10.1098/rspb.2010.1427

6. V. Iverson, R. M. Morris, C. D. Frazar, C. T. Berthiaume, R. L. Morales, E. V. Armbrust, Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota. *Science* **335**, 587–590 (2012). Medline

7. C. J. Castelle, K. C. Wrighton, B. C. Thomas, L. A. Hug, C. T. Brown, M. J. Wilkins, K. R. Frischkorn, S. G. Tringe, A. Singh, L. M. Markillie, R. C. Taylor, K. H. Williams, J. F. Banfield, Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015). Medline doi:10.1016/j.cub.2015.01.014

8. E. J. Gagen, H. Huber, T. Meador, K. U. Hinrichs, M. Thomm, Novel cultivation-based approach to understanding the miscellaneous crenarchaeotic group (MCG) archaea from sedimentary ecosystems. *Appl. Environ. Microbiol.* **79**, 6400–6406 (2013). Medline doi:10.1128/AEM.02153-13

9. J. Meng, J. Xu, D. Qin, Y. He, X. Xiao, F. Wang, Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J.* **8**, 650–659 (2014). Medline doi:10.1038/ismej.2013.174

10. C. S. Lazar, J. F. Biddle, T. B. Meador, N. Blair, K. U. Hinrichs, A. P. Teske, Environmental controls on intragroup diversity of the uncultured benthic archaea of the miscellaneous Crenarchaeotal group lineage naturally enriched in anoxic sediments of the White Oak River estuary (North Carolina, USA). *Environ. Microbiol.* **17**, 2228–2238 (2015). Medline doi:10.1111/1462-2920.12659

11. K. Kubo, K. G. Lloyd, J. F Biddle, R. Amann, A. Teske, K. Knittel, Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J.* **6**, 1949–1965 (2012). Medline doi:10.1038/ismej.2012.37

12. C. Vetriani, H. W. Jannasch, B. J. MacGregor, D. A. Stahl, A. L. Reysenbach, Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. *Appl. Environ. Microbiol.* **65**, 4375–4384 (1999). Medline

13. K. G. Lloyd, L. Schreiber, D. G. Petersen, K. U. Kjeldsen, M. A. Lever, A. D. Steen, R. Stepanauskas, M. Richter, S. Kleindienst, S. Lenk, A. Schramm, B. B. Jørgensen,

Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218 (2013). Medline doi:10.1038/nature12033

14. J. F. Biddle, J. S. Lipp, M. A. Lever, K. G. Lloyd, K. B. Sørensen, R. Anderson, H. F. Fredricks, M. Elvert, T. J. Kelly, D. P. Schrag, M. L. Sogin, J. E. Brenchley, A. Teske, C. H. House, K. U. Hinrichs, Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3846–3851 (2006). Medline doi:10.1073/pnas.0600035103

15. K. T. Konstantinidis, J. M. Tiedje, Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2567–2572 (2005). Medline doi:10.1073/pnas.0409727102

16. P. Yarza, P. Yilmaz, E. Pruesse, F. O. Glöckner, W. Ludwig, K. H. Schleifer, W. B. Whitman, J. Euzéby, R. Amann, R. Rosselló-Móra, Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645 (2014). Medline doi:10.1038/nrmicro3330

17. K. Lang, J. Schuldes, A. Klingl, A. Poehlein, R. Daniel, A. Brune, New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of "*Candidatus* methanoplasma termitum". *Appl. Environ. Microbiol.* **81**, 1338–1352 (2015). Medline doi:10.1128/AEM.03389-14

18. U. Harms, R. K. Thauer, Identification of the active site histidine in the corrinoid protein MtrA of the energy-conserving methyltransferase complex from *Methanobacterium thermoautotrophicum*. *Eur. J. Biochem.* **250**, 783–788 (1997). Medline doi:10.1111/j.1432-1033.1997.00783.x

19. W. P. Hocking, R. Stokke, I. Roalkvam, I. H. Steen, Identification of key components in the energy metabolism of the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus* by transcriptome analyses. *Front. Microbiol.* **5**, 95 (2014). 10.3389/fmicb.2014.00095 Medline doi:10.3389/fmicb.2014.00095

20. K. Ma, A. Hutchins, S. J. Sung, M. W. Adams, Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9608–9613 (1997). Medline doi:10.1073/pnas.94.18.9608

21. J. F. Biddle, S. Fitz-Gibbon, S. C. Schuster, J. E. Brenchley, C. H. House, Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10583–10588 (2008). Medline doi:10.1073/pnas.0709942105

22. U. Ermler, W. Grabarse, S. Shima, M. Goubeaud, R. K. Thauer, Crystal structure of methyl-coenzyme M reductase: The key enzyme of biological methane formation. *Science* **278**, 1457–1462 (1997). Medline doi:10.1126/science.278.5342.1457

23. D. Prakash, Y. Wu, S. J. Suh, E. C. Duin, Elucidating the process of activation of methyl-coenzyme M reductase. *J. Bacteriol.* **196**, 2491–2498 (2014). Medline doi:10.1128/JB.01658-14

24. C. T. Brown, L. A. Hug, B. C. Thomas, I. Sharon, C. J. Castelle, A. Singh, M. J. Wilkins, K. C. Wrighton, K. H. Williams, J. F. Banfield, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015). Medline doi:10.1038/nature14486

25. M. F. Haroon, S. Hu, Y. Shi, M. Imelfort, J. Keller, P. Hugenholtz, Z. Yuan, G. W. Tyson, Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**, 567–570 (2013). Medline doi:10.1038/nature12375

26. E. J. Beal, C. H. House, V. J. Orphan, Manganese- and iron-dependent marine methane oxidation. *Science* **325**, 184–187 (2009). Medline doi:10.1126/science.1169984

27. C. Rinke, P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J. F. Cheng, A. Darling, S. Malfatti, B. K. Swan, E. A. Gies, J. A. Dodsworth, B. P. Hedlund, G. Tsiamis, S. M. Sievert, W. T. Liu, J. A. Eisen, S. J. Hallam, N. C. Kyrpides, R. Stepanauskas, E. M. Rubin, P. Hugenholtz, T. Woyke, Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013). Medline

28. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015). Medline doi:10.1101/gr.186072.114

29. D. Hyatt, G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010). Medline doi:10.1186/1471-2105-11-119

30. K. A. Baublys, S. K. Hamilton, S. D. Golding, S. Vink, J. Esterle, Microbial controls on the origin and evolution of coal seam gases and production waters of the Walloon Subgroup; Surat Basin, Australia. *Int. J. Coal Geol.* **147–148**, 85–104 (2015). doi:10.1016/j.coal.2015.06.007

31. J. Bælum, C. Scheutz, J. C. Chambon, C. M. Jensen, R. P. Brochmann, P. Dennis, T. Laier, M. M. Broholm, P. L. Bjerg, P. J. Binning, C. S. Jacobsen, The impact of bioaugmentation on dechlorination kinetics and on microbial dechlorinating communities in subsurface clay till. *Environ. Pollut.* **186**, 149–157 (2014). Medline doi:10.1016/j.envpol.2013.11.013

32. R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S. M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T. W. Lam, J. Wang, SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* **1**, 18 (2012). Medline

33. D. H. Parks, *DBB: v1.0.5* (Zenodo, Geneva, 2015); https://zenodo.org/record/17426.

34. D. D. Kang, J. Froula, R. Egan, Z. Wang, MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).

35. J. Hultman, M. P. Waldrop, R. Mackelprang, M. M. David, J. McFarland, S. J. Blazewicz, J. Harden, M. R. Turetsky, A. D. McGuire, M. B. Shah, N. C. VerBerkmoes, L. H. Lee, K. Mavrommatis, J. K. Jansson, Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **521**, 208–212 (2015). Medline doi:10.1038/nature14238

36. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). Medline doi:10.1093/bioinformatics/btp324

37. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014). Medline doi:10.1093/bioinformatics/btu153

38. R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008). Medline doi:10.1186/1471-2164-9-75

39. A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, S. H. Bryant, CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015). Medline doi:10.1093/nar/gku1221

40. M. Magrane, U. Consortium, UniProt Knowledgebase: A hub of integrated protein data. *Database* **2011**, bar009 (2011). Medline doi:10.1093/database/bar009

41. K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, B. Barrell, Artemis: Sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000). Medline doi:10.1093/bioinformatics/16.10.944

42. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000). Medline doi:10.1093/nar/28.1.27

43. Y. Moriya, M. Itoh, S. Okuda, A. C. Yoshizawa, M. Kanehisa, KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007). Medline doi:10.1093/nar/gkm321

44. I. Milne, G. Stephen, M. Bayer, P. J. Cock, L. Pritchard, L. Cardle, P. D. Shaw, D. Marshall, Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**, 193–202 (2013). Medline doi:10.1093/bib/bbs012

45. T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G. L. Andersen, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006). Medline doi:10.1128/AEM.03006-05

46. K. D. Pruitt, T. Tatusova, G. R. Brown, D. R. Maglott, NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–D135 (2012). Medline doi:10.1093/nar/gkr1079

47. R. Leinonen, H. Sugawara, M. Shumway; International Nucleotide Sequence Database Collaboration, The sequence read archive. *Nucleic Acids Res.* **39**, D19–D21 (2011). Medline doi:10.1093/nar/gkq1019

48. E. R. Hawley, H. Piao, N. M. Scott, S. Malfatti, I. Pagani, M. Huntemann, A. Chen, T. Glavina Del Rio, B. Foster, A. Copeland, J. Jansson, A. Pati, S. Tringe, J. A. Gilbert, T. D. Lorenson, M. Hess, Metagenomic analysis of microbial consortium from natural crude oil that seeps into the marine ecosystem offshore Southern California. *Stand. Genomic Sci.* **9**, 1259–1274 (2014). Medline doi:10.4056/sigs.5029016

49. R. M. Soo, B. J. Woodcroft, D. H. Parks, G. W. Tyson, P. Hugenholtz, Back from the dead; the curious tale of the predatory cyanobacterium *Vampirovibrio chlorellavorus*. *PeerJ* **3**, e968 (2015). Medline doi:10.7717/peerj.968

50. R. D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy, A.

Bateman, The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222 (2010). Medline doi:10.1093/nar/gkp985

51. D. H. Haft, J. D. Selengut, O. White, The TIGRFAMs database of protein families. *Nucleic Acids Res.* **31**, 371–373 (2003). Medline doi:10.1093/nar/gkg128

52. V. M. Markowitz, I. M. Chen, K. Palaniappan, K. Chu, E. Szeto, M. Pillay, A. Ratner, J. Huang, T. Woyke, M. Huntemann, I. Anderson, K. Billis, N. Varghese, K. Mavromatis, A. Pati, N. N. Ivanova, N. C. Kyrpides, IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* **42**, D560–D567 (2014). Medline doi:10.1093/nar/gkt963

53. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009). Medline doi:10.1093/molbev/msp077

54. S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001). Medline doi:10.1093/oxfordjournals.molbev.a003851

55. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994). Medline doi:10.1007/BF00160154

56. J. Felsenstein, Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791 (1985).

57. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006). Medline doi:10.1093/molbev/msj030

58. E. P. Nawrocki, D. L. Kolbe, S. R. Eddy, Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009). Medline doi:10.1093/bioinformatics/btp157

59. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). Medline doi:10.1093/nar/gkh340

60. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011). Medline doi:10.1093/bioinformatics/btr088

61. S. Q. Le, O. Gascuel, An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008). Medline doi:10.1093/molbev/msn067

62. R. K. Bradley, A. Roberts, M. Smoot, S. Juvekar, J. Do, C. Dewey, I. Holmes, L. Pachter, Fast statistical alignment. *PLOS Comput. Biol.* **5**, e1000392 (2009). Medline doi:10.1371/journal.pcbi.1000392

63. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). Medline doi:10.1093/bioinformatics/btu033

64. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). Medline doi:10.1093/molbev/mst010

65. W. Ludwig, O. Strunk, R. Westram, L. Richter, H. Meier, A. Yadhukumar, T. Buchner, S. Lai, G. Steppi, W. Jobb, I. Förster, S. Brettske, A. W. Gerber, O. Ginhart, S. Gross, S. Grumann, R. Hermann, A. Jost, T. König, R. Liss, M. Lüssmann, B. May, B. Nonhoff, R. Reichel, A. Strehlow, N. Stamatakis, A. Stuckmann, M. Vilbig, T. Lenke, A.

Ludwig, K. H. Bode, Schleifer, ARB: A software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004). Medline doi:10.1093/nar/gkh293

66. A. Roy, A. Kucukural, Y. Zhang, I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010). Medline doi:10.1038/nprot.2010.5

67. J. Yang, A. Roy, Y. Zhang, Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **29**, 2588–2595 (2013). Medline doi:10.1093/bioinformatics/btt447

68. I. A. Berg, D. Kockelkorn, W. H. Ramos-Vera, R. F. Say, J. Zarzycki, M. Hügler, B. E. Alber, G. Fuchs, Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* **8**, 447–460 (2010). Medline doi:10.1038/nrmicro2365

69. L. Daniels, G. Fuchs, R. K. Thauer, J. G. Zeikus, Carbon monoxide oxidation by methanogenic bacteria. *J. Bacteriol.* **132**, 118–126 (1977). Medline

70. M. Rother, W. W. Metcalf, Anaerobic growth of *Methanosarcina acetivorans* C2A on carbon monoxide: An unusual way of life for a methanogenic archaeon. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16929–16934 (2004). Medline doi:10.1073/pnas.0407486101

71. J. G. Ferry, C. H. House, The stepwise evolution of early life driven by energy conservation. *Mol. Biol. Evol.* **23**, 1286–1292 (2006). Medline doi:10.1093/molbev/msk014

72. G. Borrel, N. Parisot, H. M. Harris, E. Peyretaillade, N. Gaci, W. Tottey, O. Bardot, K. Raymann, S. Gribaldo, P. Peyret, P. W. O'Toole, J. F. Brugère, Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* **15**, 679 (2014). Medline doi:10.1186/1471-2164-15-679

73. S. Heim, A. Künkel, R. K. Thauer, R. Hedderich, Thiol:fumarate reductase (Tfr) from *Methanobacterium thermoautotrophicum*: Identification of the catalytic sites for fumarate reduction and thiol oxidation. *Eur. J. Biochem.* **253**, 292–299 (1998). Medline doi:10.1046/j.1432-1327.1998.2530292.x

74. F.-P. Wang, Y. Zhang, Y. Chen, Y. He, J. Qi, K. U. Hinrichs, X. X. Zhang, X. Xiao, N. Boon, Methanotrophic archaea possessing diverging methane-oxidizing and electron-transporting pathways. *ISME J.* **8**, 1069–1078 (2014). Medline doi:10.1038/ismej.2013.212

75. T. M. Hoehler, M. J. Alperin, D. B. Albert, C. S. Martens, Apparent minimum free energy requirements for methanogenic Archaea and sulfate-reducing bacteria in an anoxic marine sediment. *FEMS Microbiol. Ecol.* **38**, 33–41 (2001). doi:10.1111/j.1574-6941.2001.tb00879.x

76. B. A. Hales, C. Edwards, D. A. Ritchie, G. Hall, R. W. Pickup, J. R. Saunders, Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Appl. Environ. Microbiol.* **62**, 668–675 (1996). Medline

77. E. Springer, M. S. Sachs, C. R. Woese, D. R. Boone, Partial gene sequences for the A subunit of methyl-coenzyme M reductase (*mcr*I) as a phylogenetic tool for the family Methanosarcinaceae. *Int. J. Syst. Evol. Microbiol.* **45**, 554–559 (1995). Medline doi:10.1099/00207713-45-3-554

78. P. E. Luton, J. M. Wayne, R. J. Sharp, P. W. Riley, The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology* **148**, 3521–3530 (2002). Medline doi:10.1099/00221287-148-11-3521

79. O. E. Håvelsrud, T. H. Haverkamp, T. Kristensen, K. S. Jakobsen, A. G. Rike, Metagenomic and geochemical characterization of pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea. *BMC Microbiol.* **12**, 203 (2012). Medline doi:10.1186/1471-2180-12-203

80. J. Bengtsson-Palme, F. Boulund, J. Fick, E. Kristiansson, D. G. Larsson, Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. *Front. Microbiol.* **5**, 648 (2014). Medline doi:10.3389/fmicb.2014.00648

81. E. Kristiansson, J. Fick, A. Janzon, R. Grabic, C. Rutgersson, B. Weijdegård, H. Söderström, D. G. Larsson, Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. *PLOS ONE* **6**, e17038 (2011). Medline doi:10.1371/journal.pone.0017038

82. L. Ma, B. Li, T. Zhang, Abundant rifampin resistance genes and significant correlations of antibiotic resistance genes and plasmids in various environments revealed by metagenomic analysis. *Appl. Microbiol. Biotechnol.* **98**, 5195–5204 (2014). Medline doi:10.1007/s00253-014-5511-3

83. T.-W. Cheng, Y. H. Chang, S. L. Tang, C. H. Tseng, P. W. Chiang, K. T. Chang, C. H. Sun, Y. G. Chen, H. C. Kuo, C. H. Wang, P. H. Chu, S. R. Song, P. L. Wang, L. H. Lin, Metabolic stratification driven by surface and subsurface interactions in a terrestrial mud volcano. *ISME J.* **6**, 2280–2290 (2012). Medline doi:10.1038/ismej.2012.61

84. Hydrocarbon Metagenomics Project, *Metagenomics for Greener Production and Extraction of Hydrocarbon Energy* (2015); www.hydrocarbonmetagenomics.com.

85. O. E. Håvelsrud, T. H. Haverkamp, T. Kristensen, K. S. Jakobsen, A. G. Rike, A metagenomic study of methanotrophic microorganisms in Coal Oil Point seep sediments. *BMC Microbiol.* **11**, 221 (2011). Medline

86. J. Kahnt, B. Buchenau, F. Mahlert, M. Krüger, S. Shima, R. K. Thauer, Post-translational modifications in the active site region of methyl-coenzyme M reductase from methanogenic and methanotrophic archaea. *FEBS J.* **274**, 4913–4921 (2007). Medline doi:10.1111/j.1742-4658.2007.06016.x

87. W. Grabarse, F. Mahlert, S. Shima, R. K. Thauer, U. Ermler, Comparison of three methyl-coenzyme M reductases from phylogenetically distant organisms: Unusual amino acid modification, conservation and adaptation. *J. Mol. Biol.* **303**, 329–344 (2000). Medline doi:10.1006/jmbi.2000.4136